Interactions of complement receptor type 2 with C3d and factor H with C3u

Thesis Presented for the Degree of

Doctor of Philosphy

by

Keying Li

Department of Structural and Molecular Biology,

University College London.

September 2010

I, Keying Li, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

Complement receptor type 2 (CR2, CD21) is a cell surface protein that links the innate and adaptive immune response through its binding to C3d, a cleavage fragment of the major complement component C3. Factor H (FH) is a major plasma protein that is the major regulator of the activity of C3b in the alternative pathway. FH binds to C3u, which is formed from C3 by hydrolysis, and C3u shows functional similarities to C3b. In this thesis, X-ray scattering, analytical ultracentrifugation and constrained modelling were used to determine solution structures and interactions of CR2 with C3d and FH with C3u. Structural studies reveal that the overall CR2 structure is unaffected by change in ionic strength or when C3d is bound to it. Unbound C3d exists in monomer-dimer and monomer-trimer equilibria in low salt buffer, but as a monomer only in physiological buffer. The CR2-C3d interaction is not formed in physiological salt conditions, but was observed in low salt conditions. The solution structure and self-association of C3u were investigated. C3u underwent weak salt-dependent dimerisation, similar to that for C3d. Modelling showed that the functionally-important TED/CUB domains in the C3d part of C3u were extended away from the rest of the C3u structure. This TED/CUB conformation is intermediate between those of C3 and C3b. C3u and FH were observed to interact as 1:1 and 2:1 complexes in a salt-dependent manner. The modelling of the interaction showed that no major conformational changes occurred in C3u or FH, and suggested that C3u binds separately to FH at two independent sites. These results provide new insights in the activation of C3 and the complement regulatory activity of CR2 and FH.

Acknowledgements

I would like to thank my supervisor Professor Stephen. J. Perkins for his guidance and encouragement throughout the period of my work with him. I am also grateful to the other two members of my thesis committee, Dr. Andrew Martin and Professor Christine Orengo, for their advice and support.

I would like to thank my collaborators, including Professor V. M. Holers and Dr. J. P. Hannan, for their expertise and for supplying the proteins and expression systems used in my studies. I am grateful for Dr Alister Dodds' advice for C3 purification.

I am very grateful to the UCL Graduate School, the Overseas Research Fund and the Mercer Fund of the Fight For Sight Charity for a postgraduate scholarship.

My thanks also go to Mr. J. Gor, Dr. A. I. Okemefuna and Miss R. Nan for their invaluable help and advice with the lab work. I thank Dr. R. George, Dr A. Bradley, Dr A. Niewiarowski and my colleagues, Ami, Nuha, Sunni and Qing at UCL for their untiring help, understanding and patience.

I would like my mum, dad, aunts and grandparents to know that I could not have done any of this without them. I would like to also thank all my friends. You've always been there for me and I really appreciated your strong backing.

Contents Page

Contents - Figures Page

Chapter One

Chapter Two

Chapter Three

Chapter Four

Chapter Five

Chapter Six

Contents - Tables

Abbreviations

Ab	antibody
AD	Alzheimer's disease
Ag	antigen
aHUS	atypical haemolytic uraemic syndrome
α2M	α2-macroglubulin
AMD	age-related macular degeneration
ANA	anaphylatoxin domain
APS	advanced photon source
AUC	analytical ultracentrifugation
AUC-FDS	analytical ultracentrifugation Fluorescence Detector System
BCR	B cell receptor
CCD	charge-coupled device
CCP	complement control protein
CD	circular dichroism
cDNA	complementary deoxyribonucleic acid
CPN	carboxypeptidase N
CRIg	complement receptor of the immunoglobulin superfamily
CRP	C-reactive protein
CR1	complement receptor type 1
CR2	complement receptor type 2
CR3	complement receptor type 3
CR4	complement receptor type 4
CUB	complement C1r/C1s, UEGF, BMP1 domain
C1INH	C1 inhibitor
C1qRp	C1q receptor
C3aR	C3a receptor
C4BP	C4b-binding protein
C5aR	C5a receptor
C5L2	C5a receptor-like protein
Da	dalton
DAF	decay-accelerating factor
Crry	complement receptor-related gene y
Crry-Ig	Crry fused to IgG1 hinge
CVF	cobra venom factor

DC-SIGN	dendritic cell-specific intercellular adhesion molecule-3-grabbing non-integrin
dIgA1	dimeric immunoglobulin A
DNA	deoxyribonucleic acid
DSSP	define secondary structure prediction
EBV	Epstein-Barr virus
E. coli	Escherichia coli
EDTA	ethylenediaminetetraacetic acid
ELISA	enzyme-linked immunosorbent assay
EM	electron microscopy
ESRF	European Synchrotron Radiation Facility
FB	factor B
Fc	fragment crystallisable
FD	factor D
FDC	follicular dendritic cell
FFT	fast Fourier transformation
FH	factor H
fHbp	factor H binding protein
FHL-1	factor H like protein-1
FHR	FH-related protein
FI	factor I
FMOD	proteoglycan fibromodulin
FT-IR	Fourier transform infrared
GAGs	glycosaminoglycans
GBM	glomerular basement membrane
GeV	giga electron volts
GM-CSF	granulocyte-macrophage colony-stimulating factor
GS	guard slits
GST	glutathione S-transferase
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfoinc acid
HIV	human immunodeficiency virus
IC	immune complex
IFN	interferon
IgA1	immunoglobulin A1
IgE	immunoglobulin E

IgG	immunoglobulin G
IgM	immunoglobulin M
IL	interleukin
ILL	Institut Laue-Langevin
K_D	dissociation constant
kb	kilobase
kDa	kilo Dalton
keV	kilo electron volts
LDL	low-density lipoprotein
LNK	linker domain
M	maximum
MAC	membrane attack complex
MAP	mitogen-activated protein
MASP	MBL-associated serine protease
MBL	mannose-binding lectin
MCP	membrane cofactor protein
mCRP	modified C-reactive protein
MG	macroglobulin
mM	millimolar
μM	micromolar
MPGN II	membranoproliferative glomerulonephritis type II
Mr	relative molecular mass
mRNA	messenger ribonucleic acid
MS	mass spectrometry
MW	molecular weight
MWCO	molecular weight cut-off
m/z	mass to charge ratio
NA	not available
NK	natural killer
NMR	nuclear magnetic resonance
PBS	phosphate-buffered saline
PDB	Protein Data Bank
PEG	polyethylene glycol
PIPES	piperazine-N, N'-bis (2-ethanesulphonic acid)
pM	picomolar

Por1A	porin 1A
ppm	part per millon
PS	primary slits
RCA	regulators of complement activation
R_G	radius of gyration
RNA	ribonucleic acid
RPE	retinal pigment epithelium
r.p.m	revolutions per minute
r.s.m.d	root mean square deviation
RU	resonance units
R_{XS}	cross-sectional radius of gyration
SANS	small angle neutron scattering
SAS	small angle scattering
SAXS	small angle X-ray scattering
SCIN	Staphylococcal complement inhibitor
SCR	short complement regulator
SDS-PAGE	sodium dodecyl suphate polyacrylamide gel electrophoresis
SE	sedimentation equilibrium
SLE	systemic lupus erythematosus
SNS	Spallation Neutron Source
SPR	surface plasmon resonance
sRPEds	sub-retinal pigment epithelial deposits
SRS	Synchrotron Radiation source
SV	sedimentation velocity
TED	thioester-containing domain
TNF	tumor necrosis factor
TS2	target station 2
TTP	thromobotic thrombocytopenic purpura
UCL	University College London
USAXS	ultra-small-angle scattering
UV	ultraviolet
VSIG4	complement receptor of the immunoglobulin superfamily
WAXS	wide-angle X-ray scattering
1D	one dimensional
2D	two dimensional

Chapter One

The complement system

(1.1) History of the complement system

Complement is a major defence and clearance system in the blood stream. The complement system plays a major role in the recognition and elimination of microbial intruders and other pathogenic cells. It is a mediator between the innate and adaptive immune responses. It consists of proteins that react to each other in a biochemical cascade and work together to eliminate invading microganisms through both antibody-dependent and independent mechanisms (Border & Gengou, 1901).

The complement system was first discovered in the late nineteenth century, when blood serum was found to lyse bacteria (Nuttal, 1888; Buchner, 1889). Then Bordet, a Belgian scientist in Paris, described the loss of the lytic activity of serum from immune animals following heating, and the subsequent restoration of activity with the addition of non-immune serum. He concluded that two factors were involved: heat-labile and heat-stable factors. He named these as "alexin" and "sensitizer" (Bordet, 1909). Alexin is later known as complement and "sensitizer" is recognised as antibody.

The term "complement" was first used by Ehrlich and Morgenroth in the late 1890s. As part of their large theory of immune system, "complement" describes a heat-labile component of the immune system that was thought to complement the action of the cells of the immune system. Ehrlich believed that each antigen-specific antibody (called amboceptor) has its own specific complement, whereas Bordet believed that there is only one type of complement. In the early 20th century, this controversy was resolved when it became understood that complement can act in combination with specific antibodies, or on its own in a non-specific way.

During the first half of the 20th century, it was established that the system comprised at least four proteins (Pillemer *et al.*, 1941). In the late 19th century, the components of complement were first observed as the globulin (also called mid-piece) and pseuglobulin (end-piece) fractions of serum. It was found that neither of them on their own could destroy bacteria. Ferrata (Ferrata, 1907) later showed that recombination of these two parts of fractions from guineas pig could destroy bacteria.

The midpiece contained a protein that could interact with erythrocyte-bound haemolytic antibody. This protein was named C1. The globulin fraction reacted with antibody first, and then the albumin fraction caused cell lysis. In the early 20th century, a third component was confirmed by several observers (Coca, 1914). On removal of this component, serum becomes inactive. In 1926, Gordon reported the existence of the fourth component of complement, showing ammonia destroyed heat-stable factor in serum, which is distinct from the third complement (Gordon *et al.,* 1926). By the late 1950s it was believed that heat-inactivated serum was devoid of C1 and C2; zymosan and cobra venom-treated serum was deficient in C3, and ammonia destroyed C4. These studies showed the sequential nature of complement activation.

With the advent of improved methods of plasma protein fractionation, most of the complement proteins became possible to purify and characterise. Mathematical models were applied to analyse the protein reactions and to clarify the steps in the sequence. By the late 1960s, complement was known to comprise eleven proteins: C1q, C1r, C1s, C4, C2, C3, C5, C6, C7, C8 and C9.

From the 1960s, the complement activation pathways were gradually discovered. In the 1960s, Nelson and Müller-Eberhard characterized an immunologically activated pathway, named as the classical pathway (Nelson *et al.*, 1966; Müller-Eberhard, 1968). Later, the alternative pathway was discovered. (Ellman *et al*., 1970; Götze & Müller-Eberhard, 1971). In 1972 Pepys (Pepys, 1972) observed an impaired acquired immune response occurs on C3-depleted mice. This finding indicated that the innate immune system is able to instruct the acquired immune response and showed that the complement system plays a more interesting role than had previously been thought. Recently, the lectin pathway was discovered with the purification of mannan-binding lectin (MBL) from rabbit liver (Kawasaki *et al*., 1978) and human (Kawasaki *et al*., 1983). Now, the each pathway is recognised as a third pathway of complement activation (Ikeda *et al*., 1987; Ji *et al*., 1988).

More than a hundred years after the initial reports, the central role of complement system in immune defence is much better known. Complement activation

activates pro-inflammatory mediators, generates anaphylactic peptides, cytolytic compounds and antimicrobial compounds, then effector cells are recruited to induce effector responses. These ensure that complement activation is moderate and controlled, in turn benefit for the host.

(1.2) Complement activation pathways

The complement system, which comprises over 30 proteins, is able to activate cells involved in both the adaptive and innate immune responses (Janeway *et al.*, 2005). The complement proteins normally circulate in the blood in an inactive zymogen form and are activated through a series of enzymatic reactions as the detection of foreign material in plasma. Complement activation can be triggered by any of following events: (a) Interaction of C1 with antigen-antibody complex, polymerised antibody IgG or IgM, or C-reactive protein (CRP), (b) spontaneous hydrolysis or molecular perturbation of the C3 complement protein, or (c) Interaction of C3 with mannan-binding lectin (MBL) that bind the microbial polysaccharide mannose.

Any of these pathways could trigger a cascade of reactions. The goal is to eliminate foreign particles from the host. This involves other complement proteins, membrane-bound control proteins and receptors, proteases and regulatory proteins. Several complement proteins are zymogens. When activated, they become proteases and activate other proteins in the complement system by cleaving them to expose reactive groups or active sites. Complement is activated by three pathways namely, the classical, the alternative and the lectin pathways (Morley & Walport, 2000). The three pathways leads to the activation of C3, and the endpoint of complement activation is the formation of the membrane attack complex (MAC).

(1.2.1) The classical pathway

Activation of complement via the classical pathway is primarily initiated by the interaction of C1q of C1 ($C1qr_2s_2$) with the Fc region of IgG and IgM antibodies. Alternatively, activation of C1 can be achieved by its direct interaction with a variety of

polyanions (such as bacterial lipopolysaccharides, DNA and RNA), certain small polysaccharides, viral membranes and C-reactive protein, etc. C1 is a complex of three proteins: C1q, C1r, and C1s. Within the C1 complex, the active sites of its subunits are positioned in close proximity, and activation of one of the subunits leads to activation of the others in turn. The C1q molecule is composed of six subunits with collagen-like tails attached to globular heads. It can bind Fc regions of antibodies IgG and IgM, which are bound to bacterial and viral surfaces. The enzymatic activity of C1 is derived by activation of the two molecules of pro-enzyme C1r and two molecules of pro-enzyme C1s in the $C1r_2$-$C1s_2$ Ca^{2+} dependent complex (Law & Reid, 1995). A model for the C1 complex is shown in Figure 1.1.

Binding of C1q to antibody causes a conformational change in C1r/C1s, which activates the protease function of C1r/C1s. C1s cleaves C4 and then C2 to generate C4b and C2b. This results in a C4b2a complex (also called C3 convertase, referring to its role of converting C3 into an active form). This convertase is attached to the surface of pathogen and cleaves C3 into C3b and C3a. The C3b molecules may diffuse away and then coat immune complexes and specific antigens, functioning as opsonin, or they bind with the C3 convertase to form a trimolecular complex C4b2a3b that is called C5 convertase. The C3b component of C5 convertase binds C5 and alters its conformation, so that the rest of component of this convertase can cleave C5 into C5a, which is a potent anaphylatoxin and diffuse away, and C5b, which attaches to C6 and induces formation of the membrane attack complex (MAC) (Figure 1.2).

(1.2.2) The alternative pathway

The alternative pathway plays an important part in the innate immune system. It does not require antibody and primarily is initiated on microbial surfaces. Plasma protein C3, factor B and factor D form a positive feedback activation loop which is controlled by the soluble regulatory components factors H, I and P (Figure 1.2). The alternative pathway is the phylogenetically oldest activation pathway of the complement system (Armstrong *et al*., 1993).

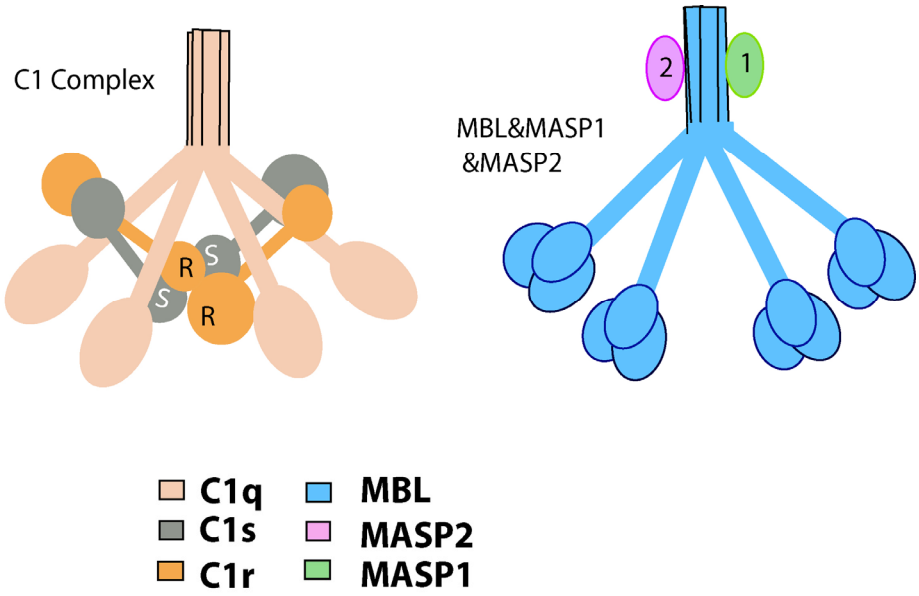


Figure 1.1 Schematic models of the C1 complex and the MBL and MASP complex. R and S denote the larger catalytic domains of C1r and C1s (Law & Reid, 1995).

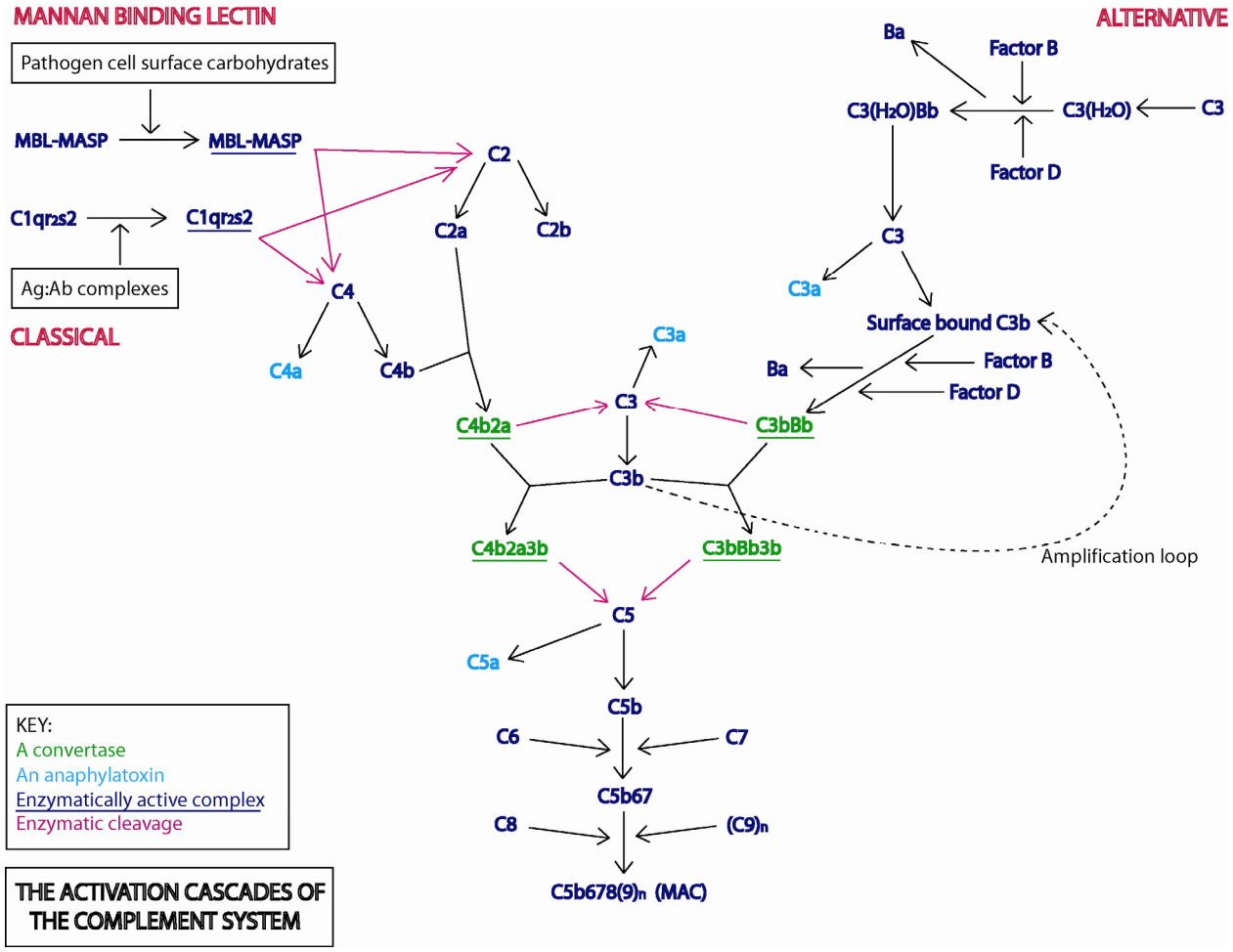


Figure 1.2 The activation cascades of the complement system (adapted from Gilbert, 2006).

C3 in the alternative pathway is activated differently from classical pathway activation. In the alternative pathway, C3, which contains an unstable thioester bond, is slowly hydrolysed to C3u or is cleaved by proteases to produce C3a and C3b. C3b binds to foreign surface antigens or even to the host cells. Due to high levels of sialic acid on most host cell surfaces which interacts with the control protein factor H, C3b is rapidly inactivated and this prevents its further interactions to lead to MAC formation. However, when a single C3b molecule is deposited on an activating surface, it can serve as a seed for the positive amplification loop which operates explosively (Nicol & Lachmann, 1973; Pangburn & Müller-Eberhard, 1978). Another serum protein called factor B interacts with C3b to form a complex. This complex is then cleaved by factor D to generate C3bBb that has C3 convertase activity and thus is analogous to the C4b2a complex in the classical pathway. The stability of C3bBb is controlled by binding to a serum protein called properdin. Because C3bBb possess enzymatic activity that cleaves unhydrolyzed C3 to generate more C3b, the initial steps are repeated and amplified. C3b also binds to C3bBb to generate the C3bBb3b complex, which exhibits C5 convertase activity, analogous to the C4b2a3b complex in the classical pathway (Law *et al.,* 1995). This is shown in Figure 1.2 also.

(1.2.3) The mannan-binding lectin pathway

Lectins are proteins that recognise and bind to specific carbohydrate targets. The lectin pathway, like the alternative pathway, is an antibody-independent pathway. However, the mechanism is very similar to the classical pathway. After initiation, it proceeds through the reactions of C2 and C4, to produce a C5 convertase; and consequently activates C5 for the formation of the membrane attack complex. The lectin pathway is activated by the binding of mannose-binding lectin (MBL) to mannose residues on glycoproteins or carbohydrates on the surface of microorganisams. MBL is an acute phase protein produced in inflammatory responses. It functions similarly to C1q to trigger complement activation (Figure 1.1). After MBL binds to the surface of a cell or pathogen, the MBL-associated serine proteases, MASP-1 and MASP-2, binds to MBL (Matsushita & Fujita, 1992). MASP-2 is an enzyme component that acts in a similar

fashion to C1s in the classical pathway, in that this is responsible for the cleavage of C2 and C4, which ultimately results in the formation of the C3 convertase enzyme C4b2a (Figure 1.2). This means of activating the C2 and C4 components to form a C5 convertase without the need for specific antibody binding represent an important innate defence mechanism comparable to the alternative pathway (Law *et al.,* 1995).

(1.2.4) Membrane attack complex

Complement activation is terminated by complement C5b, C6, C7, C8 and C9, which interact sequentially to form a supra-macromolecular structure called the membrane-attack complex (MAC). C5 is broken down into C5a and C5b after binding to C5 convertase. C5a diffuses away, and C5b is quickly degraded unless C6 binds to it to create the C5b6 complex, to which C7 then binds. The addition of C7 changes the conformation of the complex, which becomes hydrophobic and is able to insert into lipid membrane of the target cell. C8 and up to 18 molecules of C9 then bind to the complex and form the MAC (Law & Reid, 1995). C8 is a complex composed of C8α, C8β and C8α-γ, which are distinct gene products. The binding of C8 to C5b67 takes place via C8β. C8β binds to the C5b part of the membrane-bound C5b67 complex to form the C5b678 complex, and allows the hydrophobic domain of C8α-γ to be inserted into the phospholipid bilayer. The final step in formation of the MAC is binding and polymerisation of C9 to the C5b678 complex. About ten to seventeen molecules of C9 are thought to be involved. During polymerisation, the C9 molecules undergo a hydrophilic-amphilic transition, so that they can insert into the membrane. The completed MAC has a tubular form and functional pore size of 7-10 nm.

The MAC behaves as an integral membrane protein complex or transmembrane channel that is permeable to water and electrolytes. Since small molecules and ions can diffuse freely through the channel of MAC, the cell loses its osmotic stability and is consequently killed by an influx of water and loss of electrolytes (Law & Reid, 1995). Thus, complement not only facilitates the uptake and destruction of pathogens by

phagocytic cells, but also lyses cells directly and play a role in the clearance of immune complexes and inflammation.

(1.3) The regulation of complement

Complement activation occurs in a sequential manner. Once activated, the complement cascade generates effector compounds that are delivered to any surface. However, progression of the cascade and the action of the effectors need to be controlled at each level by multiple complement regulators and inhibitors. When complement activation is not properly controlled or inhibited, it can lead to inflammation, pathology and ultimately autoimmune disease.

The importance of complement regulation is highlighted by the relatively large number of regulators. Overall, active complement proteins have a short half-life, even without functional regulatory proteins. Regulators guard the correct level of activation and deactivation of complement proteins. Complement regulators are categorised in three major classes: a) in fluid phase. b) attached to the surface of host cells. c) membrane-integral complement clearance receptors. Fluid phase complement regulators are mainly in human plasma and in body fluid such as the synovial and vitreous fluids, such as factor H, properdin, Carboxypeptidase N, C1 inhibitor (C1INH), clusterin, vitronectin (Heinen *et al*., 2009; Preissner & Seiffert, 1998) and C4BP. Membrane-bound complement regulator include: CR1, CR2, CD55 (also known as DAF), CD46 (also known as MCP), CD59 (also known as protectin) and complement receptor of the immunoglobulin superfamily. Receptors for complement effectors include CR1, CR2, CR3, CR4, C3aR, C5aR, C5L2, C1qR and CRIg.

C3b, generated by C3 convertase, can damage healthy cells by causing their inappropriate opsonisation. C3b is controlled by several regulators of complement activation (RCA) proteins including Complement receptor type 1 (CR1), Membrane cofactor protein (MCP) and factor H. CR1, MCP or factor H (a regulatory component) bind to C3b thus preventing its interaction with factor B. Factor I is a serine protease

which cleaves both C3b and C4b in the presence of a cofactor (CR1, MCP or factor H). Factor H and Factor I controls the activity of the alternative pathway C3 and C5 convertases by competing with factor B for C3b binding. Factor I cleaves the C3b into a bound iC3b fragment and a soluble C3f fragment. iC3b can be further cleaved by Factor I, which releases C3c and leaves C3dg bound to the membrane. C3d, along with C3dg and iC3b, are not known to play any part in the activation of complement but are important as ligands for complement receptor type 2 (CR2) (Figure 1.3).

This thesis will focus on the regulatory interactions between CR2 and C3, and between factor H and C3. Section 1.4, 1.5 and 1.6 summarise what is known of CR2, C3 and factor H.

(1.4) Complement receptor type 2

Complement receptor type 2 (CR2/CD21) is a member of the regulators of complement activation (RCA) gene family. It is a 145 kDa membrane glycoprotein, which has an important role of cross-linking foreign antigens with surface-bound antibodies and C3d on the surface of B cells.

(1.4.1) CR2 expression and sequence

CR2 is found at the RCA gene locus (Rodriguez de Cordoba *et al.*, 1985) along with factor H, CR1, MCP, C4BP and DAF. The locus maps to human chromosome 1q32 (Weis *et al.*, 1987), in which CR2 is encoded by a single gene (Fujisaku *et al.*, 1989). Regions of internal homology have been shown in the CR2 genomic organisation (Fujisaku *et al.*, 1989) and from full length cDNA analysis (Weis *et al.*, 1988). Three types of exons were observed: (a) Two domains are encoded on the same exon, *i.e.*, short consensus repeats (SCR) domain 1 and 2, 5 and 6, 9 and 10, and 13 and 14 are encoded on single exons. (b) Exons encode a single SCR domain, where five SCRs (SCRs 3, 7, 11, 15 and 16) are encoded by single exons. (c) Three SCRs (4, 8 and 12) are encoded by split exons, where exon 4a and 4b, 8a and 8b, and 12a and 12b are split by 250,

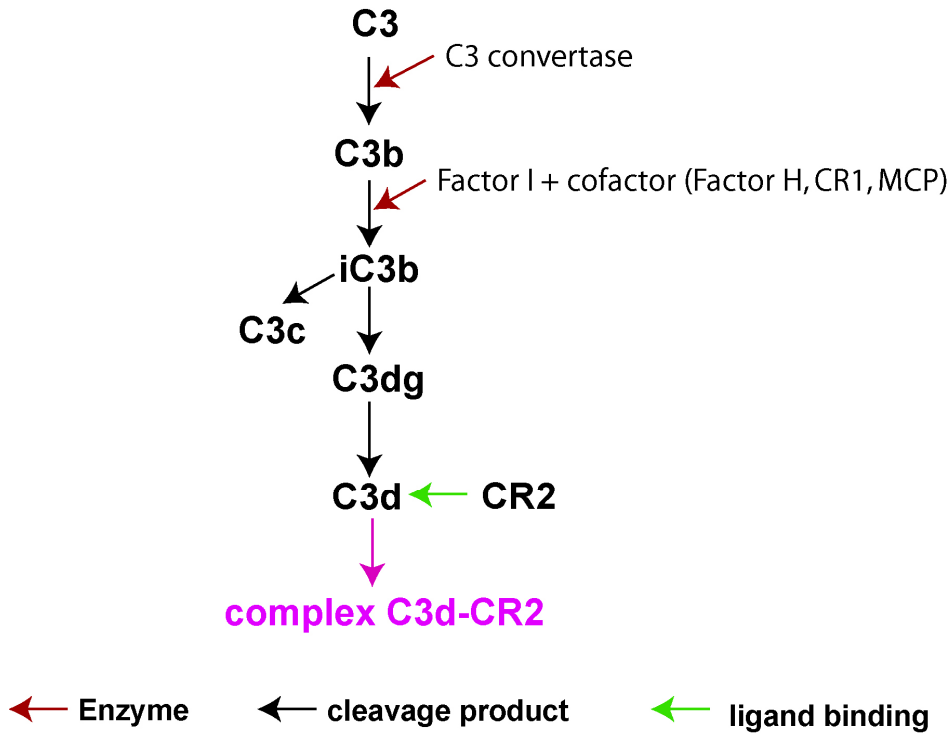


Figure 1.3 CR2 function in binding with C3 fragment C3d.

335 and 355 nucleotides respectively. The order of the exons within the gene is a repeat array of a group of four SCR domains in the order fused, single and split for the first three repeats and then fused, single, single for the last group, giving a total of 16 SCR domains (Figure 1.4). Since the fused SCR exon is present in each of the four homology group, the CR2 gene may evolved through duplication of a basic four SCR domain unit.

Two different forms of CR2 are present in cDNA libraries (Weis *et al.,* 1988; Moore *et al.,* 1987), one with 15 SCR domains, the other with 16 SCR domains, with the additional domain inserted between SCR 10 and 11 of CR2. In Figure 1.4, exon 11 occurs before 12a of 16 SCR domains. In Figure 1.5, SCR 10a is the additional domain (arrowed) in the 15 SCR domains. This is due to alternative splicing of the precursor mRNA. The intron between exon 11 and 12a in Figure 1.4 is 79 nucleotides. The efficiency of splicing to the proximal 5' splice donor site decreases when the intron is less than 80 nucleotides. (Fujisaku *et al*., 1989; Toothaker *et al.*, 1989).

CR2 is primarily expressed by B lymphocytes (Weis *et al.,* 1984), but is also found on follicular dendritic cells (Reynes *et al.,* 1985), astrocytes (Gasque *et al.,* 1996), early thymocytes (Tsoukas *et al.,* 1988), epithelial cells (Young *et al.,* 1986; Birkenbach *et al.,* 1992) and a subset of CD4+ and CD8+ peripheral T lymphocytes (Fischer *et al.,* 1991; Levy *et al.,* 1992). A soluble form of CR2 has been reported to be released spontaneously by B and T cells (Frémeaux-Bacchi *et al.,* 1996). Follicular dendritic cells selectively express the 16 SCR domain isoform of CR2 (Liu *et al.,* 1997). CR2 is expressed only in early active and mature stages of B cell development (Tedder *et al*., 1984). Expression is lost in pro-B cells, pre-B cells and in lymphocytes at the plasma cell stage.

The mature CR2 protein consists of a 111 kDa polypeptide chain with between eight and eleven putative N-linked glycosylation sites (Weis & Fearon, 1985). It consists of 15 or 16 tandem SCR domains followed by a 28 amino acid hydrophobic transmembrane domain and a relatively short 34 amino acid intracytoplasmic tail (Hannan *et al.,* 2005). Each SCR is a domain with 60-70 amino acids (Figure 1.5). Conserved cysteine and tryptophan residues are characteristic of

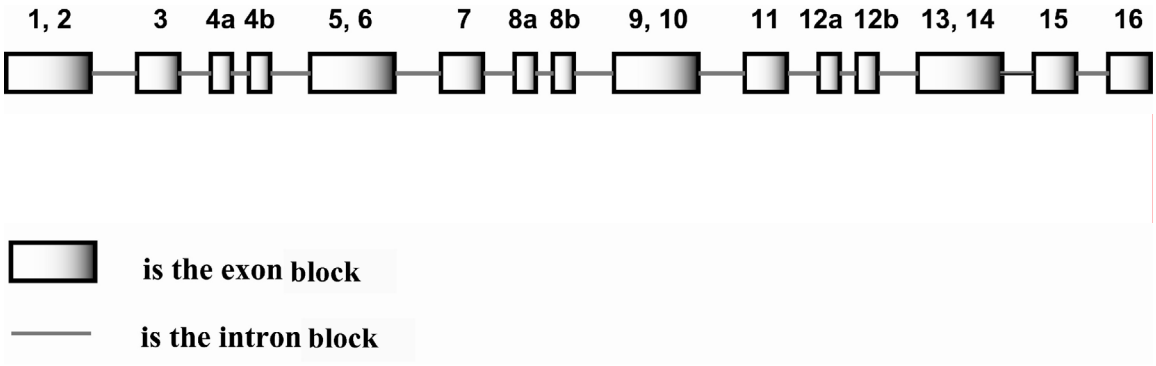
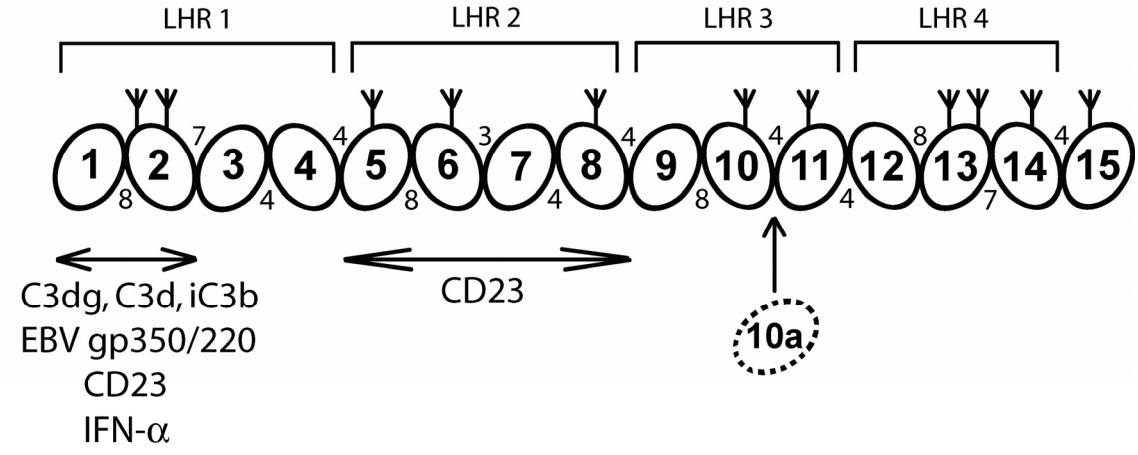


Figure 1.4 Schematic order of exons sequence within the CR2 gene, giving 16 SCR domains. The numbers above the blocks represent corresponding SCR domains which are from the derived protein sequence (Fujisaku *et al*., 1989).




```
SCR1  (21)   ISCGSPPPILNG-RISYYS-TPIAVGTVIRYSCSGT------FRLIGEKSLLCITKDKVDGTWDKP-APKCEYFNKYSS  (90)
SCR2  (91)     CPE-PIVPGGYKIRGS--TPYRHGDSVTFACKTN------FSMNGNKSVWCQA----NNMWGPTRLPTCVSVFPLE  (153)
SCR3  (154)    CPALPMIHNGHHTSENV-GSIAPGLSVTYSCESG------YLLVGEKIINCLS----SGKWSAV-PPTCEEAR    (214)
SCR4  (215)    CKSLGRFPNG-KVKEPP--ILRVGVTANFFCDEG------YRLQGPPSSRCVIAG-QGVAWTKM--PVCEEIF    (275)
SCR5  (276)    CPSPPPILNGRHIGNSL-ANVSYGSIVTYTCDPDPEEGVNFILIGESTLRCTVDSQKTGTWSGP-APRCELSTSAVQ  (350)
SCR6  (351)    CPH-PQILRGRMVSGQK-DRYTYNDTVIFACMFG------FTLKGSKQIRCNA----QGTWEPS-APVCEKE     (409)
SCR7  (410)    CQAPPNILNGQKEDRHM-VRFDPGTSIKYSCNPG------YVLVGEESIQCTS----EGVWTPP-VPQCKVAA    (470)
SCR8  (471)    CEA-----TGRQLLTKP-QHQFVRPDVNSSCGEG------YKLSGSVYQECQG--TIP--WFME-IRLCKEIT    (526)
SCR9  (527)    CPPPPVIYNGAHTGSSL-EDFPYGTTVTYTCNPGPERGVEFSLIGESTIRCTSNDQERGTWSGP-APLCKLSLLAVQ  (601)
SCR10 (602)    CSH-VHIANGYKISGKE-APYFYNDTVTFKCYSG------FTLKGSSQIRCKA----DNTWDPE-IPVCEKET    (661)
SCR11 (662)    CQH-----VRQSLQELP-AGSRVEL-VNTSCQDG------YQLTGHAYQMCQD--AENGIWFKK-IPLCKVIH    (718)
SCR12 (719)    CHPPPVIVNGKHTGMMA-ENFLYGNEVSYECDQG------FYLLGEKKLQCRSDSKGHGSWSGP-SPQCLRSPPVTR  (787)
SCR13 (788)    CPN-PEVKHGYKLNKTH-SAYSHNDIVYVDCNPG------FIMNGSRVIRCHT----DNTWVPG-VPTCIKKAFIG  (850)
SCR14 (851)    CPPPPKTPNGNHTGGNI-ARFSPGMSILYSCDQG------YLLVGEALLLCT----HEGTWSQP-APHCKEVN     (911)
SCR15 (912)    CSS-PADMDGIQKGLEPRKMYQYGAVVTLECEDG------YMLEGSPQSQCQS----DHQWNPP-LAVCRSR     (971)
```

Figure 1.5 A schematic diagram of amino acid sequence of human CR2 and its 15 SCR domains (Gilbert *et al.,* 2006b). The position of additional domain in the alternatively spliced form of CR2 is indicated by the broken ellipse labelled 10a. The bracketed ranges show the long homologous repeats (LHR). The 15th SCR does not fit the pattern. The 11 predicted N-linked glycosylation sites are marked by ψ symbols. The number of residues between each SCR domain is shown between each pair of domains. Horizontal arrows indicate the binding sites for its ligands. Underneath, the 15 SCR sequences of CR2 are aligned. (SWISSPROT entry P20023). Residues 1-20 correspond to the predicted signal peptide. Residues 21-974 contain 15 copies of the SCR sequences. The sequence from one to 20 and residue 972 onwards is not included. The N-linked glycosylation sites are underlined. Conserved cysteine and tryptophan residues are highlighted in red, and the linker sequences are highlighted in green.

the SCR structure. The presence of consensus repeating elements has been reported in other C3 binding proteins (Reid *et al*., 1986). Long homologous repeats have been reported that repeat every four SCR domains of CR2 except the last group (Weis *et al*., 1988). The cytoplasmic domain contains several potential serine and threonine phosphorylation sites (Feramisco *et al.,* 1980; Moore *et al.,* 1987; Weis *et al*., 1988). The sequence of CR2 is highly conserved across species, although the SCR 3 domain is missing in the rat CR2 (Gilbert *et al*., 2006b).

The first two SCR domains of CR2 contain binding sites for four known ligands (Figure 1.5). These include the C3 fragments (C3d, C3dg, and iC3b) (Lowell *et al.,* 1989), the Epstein-Barr virus (EBV) coat glycoprotein gp 350/220 (Carel *et al.,* 1990), interferon alpha, IFNα (Delcayre *et al.,* 1991) and CD23 (Aubry *et al.,* 1992). The SCR 5-8 domains also have a binding site for CD23 (Aubry *et al.,* 1994).

(1.4.2) CR2 function

CR2 plays an important role in the generation of a normal immune response. Its deficiency causes an impaired B cell response to T-dependent antigen (Ahearn *et al.,* 1996). For mature B lymphocytes, CR2 acts as a co-receptor for IgM and the B cell receptor (BCR). Cross-linking of CR2 and membrane IgM can cause a synergistic increase in the free intracellular calcium concentration in human B lymphocytes (Carter *et al*., 1988). It enhances the response of B cells to antibodies and subsequently enlarging the T cell responses (Hebell *et al.,* 1991; Thyphronitis *et al.,* 1991; Ahearn *et al.,* 1996). In this way CR2 acts as a linker between the innate and humoral acquired immune response (Dempsey *et al.,* 1996).

CR2 mediates Epstein-Barr virus (EBV) infection by serving as the receptor for the EBV intrinsic membrane protein gp350/220 (Fingeroth *et al.,* 1984). EBV is a human herpes virus which is carried in a latent state in the B cells of approximately 90% of humans. In most cases there are no symptoms, but EBV can cause infectious mononucleosis and is associated with Hodgkin's lymphoma, Burkitt's lymphoma and adult T cell leukaemia (Epstein *et al*., 1964; Weiss *et al*., 1989; Jones *et al*., 1988). EBV binds to CR2 on B cells through its coat glycoprotein gp350. The binding site on CR2 for gp350 has

been shown in the SCR1-2 domains by site-mutagenesis (Young *et al*., 2008). The discovery of this binding site may assist investigation of EBV therapy.

CR2 was assessed for its importance for the binding of the human immunodeficiency virus immune complex (HIV-IC). Previous study showed that CR2 is a critical receptor for high-level HIV IC binding to PBMC and lymph node mononuclear cells (Jakubik *et al*., 1999). The C3 fragments are deposited on HIV-1 and bind to CR1 and CR2 on B cells. As the B cells pass through the secondary lymphoid organs, the virions may be passed to FDC and subsequently to T cells. The discovery may assist investigation on HIV therapy.

CR2 has four known types of ligands: C3 cleavage fragments, CD23 (a low affinity receptor for IgE), IFNα and the coat glycoprotein gp350 of EBV (Makar *et al.,* 1998). Research has revealed that the SCR1-2 domains contain binding sites for C3dg, iC3b, gp350 and CD23 (CD23 also interacts with CR2 SCR5-8). CR2 on B cells provides a preferential site for C3 fragments with covalently attached antigen. The cross-linking of CR2 to the B-cell receptor by a C3d-antigen complex enhanced an intracellular calcium release through a co-activation with CD19 and CD81 (Fearon & Carroll, 2000). Also, CR2 is a receptor for CD23. CD23, which has been shown to influence B cell growth and proliferation and germinal B cell survival (Liu *et al.,* 1991), is a type two transmembrane protein that is a low affinity receptor for IgE. It interacts with CR2 at both SCR1-2 and SCR5-8. CD23 can interact with CR2 to increase interleukin-4 induced production of IgE.

(1.4.3) Protein structure of CR2

Since CR2 plays an important part in the human complement system, a structural understanding of it becomes essential. Hydrodynamic and EM studies of soluble CR2 showed that it is a highly extended, rod-shape and flexible molecule (Moore *et al.,* 1989). Figure 1.6 shows that CR2 has a beads-on-a-string arrangement and a length of 38.6 ± 3.5 nm overall. The crystal structure of CR2 SCR1-2 shows the first two domains from N- terminal form a closed V shape (Prota *et al*., 2002). Later, the structure of CR2 SCR1-15 in solution was determined by constrained scattering and ultracentrifugation modelling (Gilbert *et al*., 2006b). The length of the molecule is estimated as 38 nm. This is consistent with electron microscopy images with 16 SCR domains. The best-fit CR2 models in solution are extended

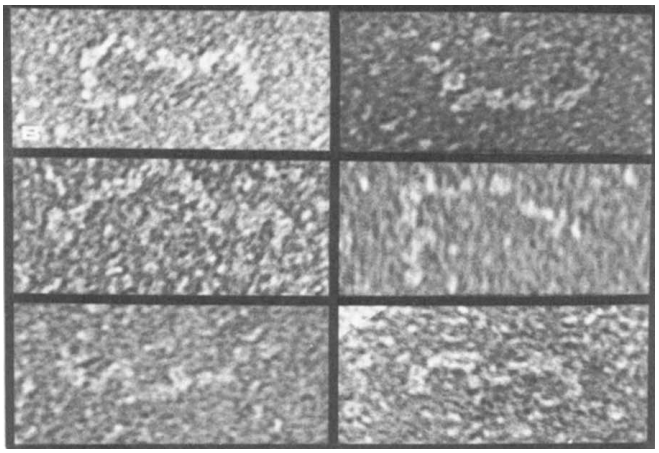


Figure 1.6 Structures of CR2 by electron microscopy. Sixteen extracellular domains of CR2 are observed here (taken from Moore *et al.,* 1989).

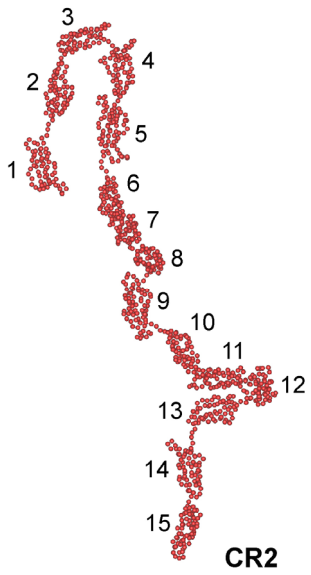


Figure 1.7 CR2 SCR1-15 conformation with 15 SCR domains by solution scattering (adapted from Gilbert *et al.,* 2006b).

and flexible with folded back regions. As shown in Figure 1.7, SCR1-2 folds back towards SCR 4-6, and with a further slight turn around SCR 11-13; however, due to the medium resolution of solution scattering, other partially folded-back structures are also possible. The folded back regions are a particular feature of the SCR superfamily. Previous solution studies for proteins with five SCR domains (mouse Crry-Ig, rat Crry and β2-glycoprotein I) display no tendency to form such significantly folded back SCR structures (Aslam *et al.,* 2003; Hammel *et al.,* 2002), while these have been observed for factor H (Aslam & Perkins, 2001).

There are several possibilities for causes of folded back SCR regions. To date, folded-back SCR regions all possess longer than usual linkers of lengths seven or eight residues in place of three or four. This may be caused by weak interaction between the SCR domains in distant part of the molecule. A Poisson–Boltzmann electrostatic map suggested that SCR 1-2 have predominantly basic surfaces, while SCR 5 has a partial acidic surface (mediated by Asp304, Asp306, Glu308 and Glu309). These could lead to an attractive force for folding back between non-neighbouring SCR domains. Protein–carbohydrate interactions could also cause protein folding back, but it is unlikely to occur, because the interactions are transient in relation to the observed extended CR2 structures. If the proposed inter-SCR interactions were stronger, it would lead to the formation of CR2 dimers in solution; however, there is no evidence of CR2 oligomer from previous scattering and ultracentrifugation work (Moore *et al.,* 1989; Gilbert *et al.,* 2005; Gilbert *et al.,* 2006b).

The flexibility of CR2 SCR1-15 reveals the possibility for all 15 domains playing a role in ligand binding. Figure 1.5 shows the possible binding region for gp350, CD23, iC3b, IFN-α and C3d on CR2. Previous affinity studies of the CR2 ligands, C3d, IFN-α and gp350, which bind to domain SCR1-2, SCR1-4 and SCR1-15, showed an increase in dissociation constant (K_D) for the complex as the number of SCR domains increased from two to 15 (Asokan *et al.,* 2006). This may be due to a movement of the SCR 1-2 domains closer to SCR 3-15 to obstruct the ligand binding site at SCR 1-2 (Figure 1.7). Another scenario is that SCR 1-2 may interact weakly with other SCR domains in CR2. In such a case, the CR2 ligands will compete with other SCR domains for binding to SCR 1-2, and this could also explain the reduced ligand affinity of CR2 SCR 1-15 compared to CR2 SCR 1-2. Experiments with a series of truncated CR2 constructs showed that SCR 1-2 are required for ligand binding,

SCR 3-4 increase the binding affinity, and the remaining SCR domains mediate the internalisation of the bound ligand (Carel *et al.,* 1990).

The primary ligand for CR2 is the C3d activation fragment of complement component C3. The binding of C3d to CR2 on mature B lymphocytes was said to cause reduction in the threshold of immune activation. C3d has been reported to show a self-association of monomer-dimer in 50 mM NaCl, but this disappears during its binding with CR2 (Gilbert *et al*., 2005). The crystal structure of the complex CR2 SCR1-2 and C3d showed that only the SCR2 domains of CR2 interacted with C3d (Szakonyi *et al*., 2001). Mutation studies on complex confirmed the ligand binding to SCR2 (Hannan *et al*., 2005). By point mutation, SCR1 domain was also found to participate in C3d binding; in particular four residues of SCR1 affect ligand binding- Arg13, Arg28, Lys41 and Lys57 (labelled in Figure 1.8). The solution structure of CR2 SCR 1-2, both free and complexed with C3d, is a much more open V-shape than seen in the crystal structures (Figure 1.9(b,c)) (Gilbert *et al.,* 2005). This difference showed a high degree of flexibility in the linker between SCR 1 and 2, which is able to adopt a wide range of different conformations with its ligands. Electrostatic calculations have been reported for the CR2-C3d complex (Hannan *et al*., 2005; Morikos & Lambris, 2004; Zhang *et al*., 2007; Young *et al*., 2007). These calculations show that C3d has an excess charge of -4, while CR2 SCR1-2 has an excess charge of +8, and that CR2 SCR1-2 shows a preponderance of positive charge on its surface (Figure 1.8). The calculations postulate that electrostatic interactions play a role in complex formation.

What is not clear from these studies is the stoichiometry of the CR2-C3d interaction in solution. Surface plasmon resonance studies in 50 mM NaCl HEPES buffer for CR2 SCR1-2/C3d, CR2 SCR1-4/C3d and CR2 SCR1-15/C3d demonstrate a high affinity for C3d, but in more physiological 125 mM NaCl buffer, the binding is much weakened (Asokan *et al*., 2006; Moore *et al*., 1989; Guthridge *et al*., 2001; Hannan *et al*., 2005). It is known that human physiological fluid contains 137 mM NaCl, and CR2-C3d binding must happen to trigger further signal transduction. While it is assumed that the binding affinity will be weakened in higher salt, it is not known whether association of CR2 and C3d is affected in vivo. A second question is whether CR2 undergo conformational change from 50 mM NaCl to 137 mM NaCl similar to other flexible SCR proteins, such as factor H

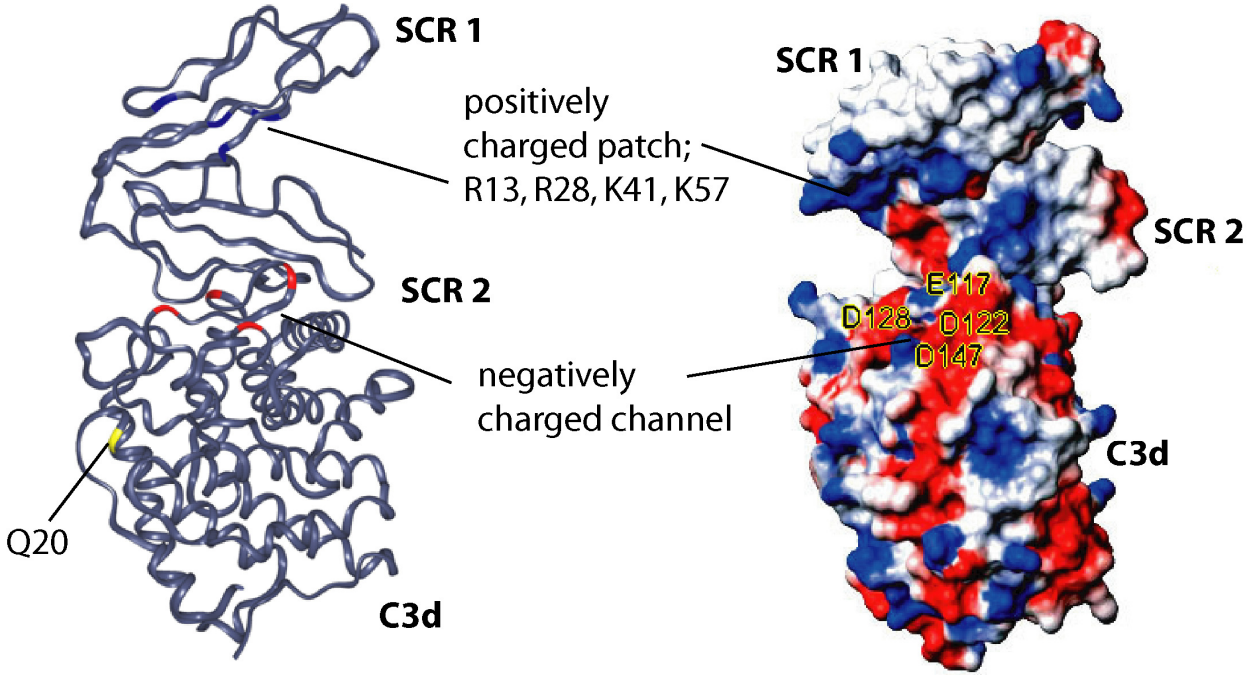


Figure 1.8 Electrostatic representation of the CR2 SCR1-2/C3d crystal structure (Hannah *et al*., 2005; Gilbert, 2006). Residues on C3d E117, D122, D128 and D147 form part of a conserved negatively charged groove located on C3d. On the left side, they are coloured red and labelled in black/yellow font on the right side. The contact residue Q20 on C3d for antigen is labelled and coloured yellow. The contacting surface of CR2 SCR2 is mainly positive charged (blue). Positive charged residues R13, R28, K41 and K57 were also found on SCR2 domain of CR2.

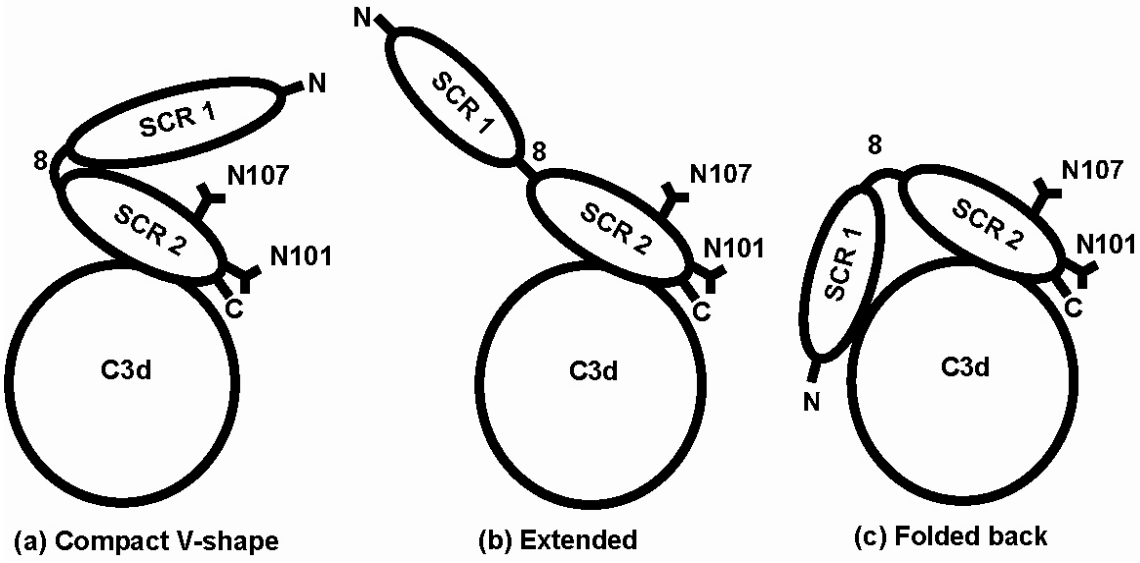


Figure 1.9 Schematic diagrams of the CR2 SCR 1-2/C3d complex to show three alternative models for the complex. The crystal structure corresponds to (a), while the solution scattering structures corresponds to (b) and (c) (adapted from Gilbert *et al*., 2005).

(Okemefuna *et al*., 2009a). NMR or crystallography is not able to analyse the structure of full length CR2 SCR1-15 and the interaction between full length CR2 SCR1-15 and C3d; therefore, other structural methods are needed to study the overall structure. Chapter 4 reports studies of the interaction between CR2 SCR1-15 and C3d by solution scattering and analytical ultracentrifugation in 50 mM NaCl and 137 mM NaCl buffers. The results provide further key clarification of the solution structure of the CR2-C3d complex and its involvement with the BCR as a coreceptor that links innate and adaptive immunity.

(1.5) Complement component C3

C3d is a binding partner of CR2. C3d originates as a fragment of the major complement component C3. C3 is the most abundant complement component in serum. It is present in plasma at a concentration of about 1.0 mg/ml, but this level can be elevated considerably during inflammation and infection. The predominant producer of C3 is hepatocytes (Alper *et al*., 1969; Morris *et al*., 1982). A number of extrahepatic cells also produce C3, including leukocytes, fibroblasts, keratinocytes and endothelial cells (Basset-Seguin *et al*., 1990; Ueki 1987).

(1.5.1) C3 expression and biosynthesis

Human C3 is coded for by a 41 kb gene located on the distal long arm on chromosome 19. The C3 gene is organised in 41 exons. The primary structure, deduced from the cDNA sequence, has 1663 amino acid residues, including a 22 amino acid signal peptide. The protein is synthesised as a single chain pre-pro molecule with the α-chain (992 residues) and β-chain (645 residues) linked by a tetra-arginine sequence, which is removed by an enzyme during posttranslational modification. The α- and β-chains, linked by a disulphide bond and noncovalent forces, have molecular weight of 115 and 75 kDa respectively. Following translocation through endoplasmic reticulum to the Golgi, three N-linked carbohydrate moieties are attached to the α- and β-chain (Perkins & Sim, 1986; Hase *et al*., 1985; Hirani *et al*., 1986; Section 1.5.3). Native C3 is a 189 kDa glycoprotein that contains three high-mannose type carbohydrates at Asn63, Asn917 and Asn1617. (Hirani *et al*., 1986; Miki *et al*., 1986; Hase *et al*., 1985). The α-chain carbohydrate may contain $Man_8GlcNAc_2$ or $Man_9GlcNAc_2$, while the β-chain may have $Man_5GlcNAc_2$ and $Man_6GlcNAc_2$. The carbohydrates together form about 2.3% of the molecular weight of C3.

In human, based on the agrose gel electrophoretic mobility of the allotypes, two major C3 polymorphisms have been identified (Welch *et al*., 1990, 1993), namely the C3 slow (C3S) and C3 fast (C3F) variants (R102G , N1216D in allele C3F; Figure 1.10). C3S is the more common allotype, with a frequency of 0.79 and 0.99 in White and Asian populations. Welch's study showed that C3F had a lower activity than C3S in cell-surface binding. Bartók and Walport found no differences between binding of C3S and C3F to the majority of CR1, CR2 and CR3. (Bartók & Walport, 1995). A recent study showed the functional difference between C3S and C3F. In recipients of kidney transplantation who received C3F/F or C3F/S donor allotype graft, survival was significantly better than a C3S/S allotype (Brown *et al*., 2006). C3F is also related to age-related macular degeneration (Yates *et al*., 2007).

Likewise many other acute-phase reactants, the synthesis of C3 by hepatocytes is regulated by tumour necrosis factor, interleukin (IL)-1 β and interleukin-6. In addition, the extrahepatic C3 synthesis is regulated by cytokines of different cellular origin. Most of the cytokines up-regulate C3 expression and synthesis, but granulocyte-macrophage colony-stimulating factor (GM-CSF) and interferons inhibit basal C3 production in monocytes. Glucocorticoids and aldosterone also increase production of C3 in alveolar epithelial cells. As for rat, C3 produced by endometrium is regulated by oestrogen and progesterone (Hasty & Lyttle, 1992; Puy *et al*., 1993).

(1.5.2) C3 sequence and structure

Human C3 has 79% sequence identity to mouse C3 at the nucleotide level and 77% sequence identity at the amino acid level. Of the codons, 51% are unchanged and 26% have changed conservatively at the amino acid level. The similarity is highest in the α chain and lowest for the signal peptide (Figure 1.10).

C3 is a member of the α2-macroglobulin (α2M) family of proteins. The proteins of this family have a primary function in host defense (Blandin & Levashina, 2004). All these proteins are relatively large (1400-1800 aa) with several marked functional features; a reactive thioester moiety enabling covalent attachment to target particles, and a highly variable central segment likely involved in recognition (Budd *et al*., 2004). C3, C4 and α2-macroglobulin belong to the family (Reid & Porter, 1981). The protease α2-macroglobulin and complement component C4 show considerable homology to C3. The protein pairs of

```
                                        >------ MG1 ------------------------
C3_mouse      MGPASGSQLLVLLLLLASSPLALGIPMYSIITPNVLRLESEETIVLEAHDAQGDIPVTVT 60
C3_rat        MGPTSGSQLLVLLLLLASSLLALGSPMYSIITPNVLRLESEETFILEAHDAQGDVPVTVT 60
C3_cavpo      MGPAAGP--SLLLLLLASVSLALGDPMYSIITPNILRLENEETVVLEAHEVQGDIPVTVT 58
C3_bovin      MKPTSGP--SLLLLLLASLPMALGNPMYSMITPNILRLESEETVVLEAHGGQGTIQVSVT 58
C3_human      MGPTSGP--SLLLLLLTHLPLALGSPMYSIITPNILRLESEETMVLEAHDAQGDVPVTVT 58
              * *::*.   :*****:   :*** ****:****:****.***.:****  ** : *:**

              ---------------------------- MG1 ----------------------------
C3_mouse      VQDF-LKRQVLTSEKTVLTGASGHLRSVSIKIPASKEFNSDKEGHKYVTVVANFGETVVE 119
C3_rat        VQDF-LKKQVLTSEKTVLTGATGHLNRVFIKIPASKEFNADK-GHKYVTVVANFGATVVE 118
C3_cavpo      VHDFPAKKNVLSSEKTVLTSATGYLGTVTIKIPASKEFKSDK-GRKLVVVQAAFGGTQLE 117
C3_bovin      VHDFPAKKQVLSNENTQLNSNNGYLSTVTIKIPASKELKSDK-GHKFVTVVATFGNVQVE 117
C3_human      VHDFPGKKLVLSSEKTVLTPATNHMGNVTFTIPANREFKSEKGRNKFVTVQATFGTQVVE 118
              *:**  *: **:.*:* *.  ..::  * :.***.:*::::*  .* *.* * **   :*

              ------>------------------- MG2 ------------------------------
C3_mouse      KAVMVSFQSGYLFIQTDKTIYTPGSTVLYRIFTVDNNLLPVGKTVVILIETPDGIPVKRD 179
C3_rat        KAVLVSFQSGYLFIQTDKTIYTPGSTVFYRIFTVDNNLLPVGKTVVIVIETPDGVPIKRD 178
C3_cavpo      KVVLVSLQSGYLFIQTDKTIYTPGSTVLYRIFTVDSDLLPVGRTIIVTIETPDGIPIKRD 177
C3_bovin      KVVLISLQSGYLFIQTDKTIYTPGSTVLYRVFTVDHKLLPVGQTVFITIETPDGIPVKRD 177
C3_human      KVVLVSLQSGYLFIQTDKTIYTPGSTVLYRIFTVNHKLLPVGRTVMVNIENPEGIPVKQD 178
              *.*::*:********************:**:***: .*****:*:.: **.*:*:*:*:*

              ---------------------------- MG2 ---------------->----------
C3_mouse      ILSSNNQHGILPLSWNIPELVNMGQWKIRAFYEHAPKQIFSAEFEVKEYVLPSFEVRVEP 239
C3_rat        ILSSHNQYGILPLSWNIPELVNMGQWKIRAFYEHAPKQTFSAEFEVKEYVLPSFEVLVEP 238
C3_cavpo      TLSSNNQHGILPLSWNIPELVNMGQWKIQAFYENSPKQVFSAEFEVKEYVLPSFEVLVEP 237
C3_bovin      SKSSQNQFGILTLSWNIPELVNMGVWKIKAYYEDSPQQVFSAEFEVKEYVLPSFEVQLEP 237
C3_human      SLSSQNQLGVLPLSWDIPELVNMGQWKIRAYYENSPQQVFSTEFEVKEYVLPSFEVIVEP 238
                **:** *:*.***:******** ***:*:**.:*:* **:*************** :**

              ---------------------------- MG3 ----------------------------
C3_mouse      TETFYYIDDPNGLEVSIIAKFLYGKNVDGTAFVIFGVQDGDKKISLAHSLTRVVIEDGVG 299
C3_rat        TEKFYYIHGPKGLEVSITARFLYGKNVDGTAFVIFGVQDEDKKISLALSLTRVLIEDGSG 298
C3_cavpo      TEKFYYIDDPKGLEVNIIARFLYGKNVDGTAFVIFGVQDGDQRISLAQSLTRVVIEDGSG 297
C3_bovin      EEKFYYIDDPDGLKVNIIARFLYGEQVDGTAFVIFGVQDGDRRISLTHSLTRVPINDGNG 297
C3_human      TEKFYYIYNEKGLEVTITARFLYGKKVEGTAFVIFGIQDGEQRISLPESLKRIPIEDGSG 298
                *.**** . .**:*.* *:****::*:********:** :::***. **.*: *:** *
              ---------------------------- MG3 ------------------->--------
C3_mouse      DAVLTRKVLMEGVRPSNADALVGKSLYVSVTVILHSGSDMVEAERSGIPIVTSPYQIHFT 359
C3_rat        EAVLSRKVLMDGVRPSSPEALVGKSLYVSVTVILHSGSDMVEAERSGIPIVTSPYQIHFT 358
C3_cavpo      EVVLSRQVLLDGVQPSRPEALVGKSLYVSVTVILHSGSDMVEAERSGIPIVTSPYQIHFT 357
C3_bovin      EAILKRQVLLNGVQPSRADALVGKSIYVSATVILQSGSDMVEAERTGIPIVTSPYQIHFT 357
C3_human      EVVLSRKVLLDGVQNPRAEDLVGKSLYVSATVILHSGSDMVQAERSGIPIVTSPYQIHFT 358
              :.:*.*:**::**: . .: *****:***.****:******:***:**************

              ---------------------------- MG4 ----------------------------
C3_mouse      KTPKFFKPAMPFDLMVFVTNPDGSPASKVLVVTQGSN-AKALTQDDGVAKLSINTPNSRQ 418
C3_rat        KTPKFFKPAMPFDLMVFVTNPDGSPARRVPVVTQGSD-AQALTQDDGVAKLSVNTPNNRQ 417
C3_cavpo      KTPKYFKPAMPFEIMVLVTNPDGSPAPHVPVVTQGSN-VQSLTQADGVARLSINTPNTRQ 416
C3_bovin      KTPKFFKPAMPFDLMVYVTNPDGSPARHIPVVTQGSN-VQSLTQDDGVAKLSINTQNKRD 416
C3_human      KTPKYFKPGMPFDLMVFVTNPDGSPAYRVPVAVQGEDTVQSLTQGDGVAKLSINTHPSQK 418
              ****:***.***::** ********* :: *..**.: .::*** ****:**:**  .:.

              ---------- MG4 ------------>------------ MG5 -------------
C3_mouse      PLTITVRTKKDTLPESRQATKTMEAHPYSTMHNSNNYLHLSVSRMELKPGDNLNVNFHLR 478
C3_rat        PLTITVSTKKEGIPDARQATRTMQAQPYSTMHNSNNYLHLSVSRVELKPGDNLNVNFHLR 477
C3_cavpo      PLSVTVQTKKGGIPDARQAINTMQALPYTTMYNSNNYLHLSMPRTELKPGETINVNFHLR 476
C3_bovin      PLTITVRTKKDNIPEGRQATRTMQALPYNTQGNSNNYLHLSVPRVELKPGETLNVNFHLR 476
C3_human      PLSITVRTKKQELSEAEQATRTMQALPYSTVGNSNNYLHLSVLRTELRPGETLNVNFLLR 478
              **::** ***  :.:..** .**:* **.*  *********: * **:**:.:**** **

              ---------------------------------- MG5 ----------------------
C3_mouse      TDPGHEAKIRYYTYLVMNKGKLLKAGRQVREPGQDLVVLSLPITPEFIPSFRLVAYYTLI 538
C3_rat        TDAGQEAKIRYYTYLVMNKGKLLKAGRQVREPGQDLVVLSLPITPEFIPSFRLVAYYTLI 537
```

```
C3_cavpo      SDPNQEAKIRYYTYLIMNKGKLLKVGRQPREPGQALVVLPMPITKELIPSFRLVAYYTLI 536
C3_bovin      TDPGEQAKIRYYTYMIMNKGKLLKVGRQYREPGQDLVVLPLTITSDFIPSFRLVAYYTLI 536
C3_human      MDRAHEAKIRYYTYLIMNKGRLLKAGRQVREPGQDLVVLPLSITTDFIPSFRLVAYYTLI 538
                *  .:********::****:***.*** ***** ****.:.** ::**************

                ------ MG5 -----→------------------- MG6 ------------------
C3_mouse      GASGQREVVADSVWVDVKDSCIGTLVVKG----DPRDN--HLAPGQQTTLRIEGNQGARV 592
C3_rat        GANGQREVVADSVWVDVKDSCVGTLVVKG----DPRDNR-QPAPGHQTTLRIEGNQGARV 592
C3_cavpo      GASAQREVVADSVWADVRDSCVGTLVVKGGSGKDGQDKRQQHLPRQQMTLRIEGNQGARV 596
C3_bovin      NAKGQREVVADSVWVDVKDSCMGTLVVKN----GGKEEK-HHRPGQQITLKIEADQGARV 591
C3_human      GASGQREVVADSVWVDVKDSCVGSLVVKS-----GQSEDRQPVPGQQMTLKIEGDHGARV 593
              .*..**********.**:***:*:****.      :.:  :  * :* **:**.::****

              -MG6->----------------- LNK --------------------------------
C3_mouse      GLVAVDKGVFVLNKKNKLTQSKIWDVVEKADIGCTPGSGKNYAGVFMDAGLAFKTS-QGL 651
C3_rat        GLVAVDKGVFVLNKKNKLTQSKIWDVVEKADIGCTPGSGKNYAGVFMDAGLTFKTN-QGL 651
C3_cavpo      GLVAVDKGVFVLNKKHKLTQSKIWDVVEKADIGCTPGSGKDYAGVFTDAGLSFKSSKAGL 656
C3_bovin      GLVAVDKGVFVLNKKNKLTQRKIWDVVEKADIGCTPGSGRNYAGVFTDAGLTLKTS-QGL 650
C3_human      VLVAVDKGVFVLNKKNKLTQSKIWDVVEKADIGCTPGSGKDYAGVFSDAGLTFTSS-SGQ 652
                **************:**** ******************::***** ****::.:.  *

              --- LNK -->--------------------- ANA ----------------------
C3_mouse      QTEQRADLECTKPAARRRRSVQLMERRMDKAGQYTDKGLRKCCEDGMRDIPMRYSCQRRA 711
C3_rat        QTDQREDPECAKPAARRRRSVQLMERRMDKAGQYTDKGLRKCCEDGMRDIPMPYSCQRRA 711
C3_cavpo      QTAQREGLDCPKPAARRRRSVQLMERRMDKAGKYKSKELRRCCEDGMRENPMQFSCQRRA 716
C3_bovin      ETQQRADPQCPQPATRRRRSVQLMEKRMDKAGQYSS-DLRKCCEDGMRDNPMKFPCQRRA 709
C3_human      QTAQRAELQCPQPAARRRRSVQLTEKRMDKVGKYPK-ELRKCCEDGMRENPMRFSCQRRT 711
              :* **   :*.:**:******** *:****.*:* .  **:*******: ** :.****:

              ---------------- ANA ----------------→------- α'NT --→----
C3_mouse      RLITQGENCIKAFIDCCNHITKLREQHRRDHVLGLARSELEEDIIPEEDIISRSHFPQSW 771
C3_rat        RLITQGESCLKAFMDCCNYITKLREQHRRDHVLGLARSDVDEDIIPEEDIISRSHFPESW 771
C3_cavpo      RYVSLGEACVKAFLDCCTYMAQLRQQHRREQNLGLARSDMDEDIIPEEDIISRSQFPESW 776
C3_bovin      QFILQGDACVKAFLDCCEYITQLRQQHSRDGALELARSDLDDDIIPEEDIISRSQFPESW 769
C3_human      RFISLGEACKKVFLDCCNYITELRRQHARASHLGLARSNLDEDIIAEENIVSRSEFPESW 771
              : :  *: * *.*:*** ::::**.** *   * ****:::***.**:*:***.**:**

              ---------------------------- MG6 ---------------------→----
C3_mouse      LWT-IEELKEPEKNGISTKVMNIFLKDSITTWEILAVSLSDKKGICVADPYEIRVMQDFF 830
C3_rat        LWT-IEELKEPEKNGISTKVMNIFLKDSITTWEILAVSLSDKKGICVADPYEITVMQDFF 830
C3_cavpo      LWT-IEELKEPERNGISTKTMNIFLKDSITTWEILAVSLSDKKGICVADPFEVTVMQDFF 835
C3_bovin      LWTVIEDLKQADKNGISTKLMNVFLKDSITTWEILAVSLSDKKGICVADPYEVTVMQDFF 829
C3_human      LWN-VEDLKEPPKNGISTKLMNIFLKDSITTWEILAVSMSDKKGICVADPFEVTVMQDFF 830
              **. :*:**:. :****** **:***************:***********:*: ******

              ------------------------- MG7 ----------------------------
C3_mouse      IDLRLPYSVVRNEQVEIRAVLFNYREQEELKVRVELLHNPAFCSMATAKNRYFQTIKIPP 890
C3_rat        IDLRLPYSVVRNEQVEIRAVLFNYREQEKLKVRVELLHNPAFCSMATAKKRYYQTIEIPP 890
C3_cavpo      IDLRLPYSVVRNEQVEIRAVLYNYREAQSLKVRVELLHNPAFCSLATAKKRHTQTVTIGP 895
C3_bovin      IDLRLPYSVVRNEQVEIRAILYNYREAENLKVRVELLYNPAFCSLATAKKRHQQTITIPA 889
C3_human      IDLRLPYSVVRNEQVEIRAVLYNYRQNQELKVRVELLHNPAFCSLATTKRRHQQTVTIPP 890
              ********************:*:***: :.********:******:**:*.*: **: * .

              ------------ MG7 ------------------------→------- CUB ----
C3_mouse      KSSVAVPYVIVPLKIGQQEVEVKAAVFNHFISDGVKKTLKVVPEGMRINKTVAIHTLDPE 950
C3_rat        KSSVAVPYVIVPLKIGLQEVEVKAAVFNHFISDGVKKILKVVPEGMRVNKTVAVRTLDPE 950
C3_cavpo      KSSVAVPYVLVPLKIGLQEVEVKAAVYNYFISDGVKKTLKVVPEGMRVNKTVAIRTLNPE 955
C3_bovin      RSSVAVPYVIVPLKIGLHEVEVKAAVYNHFISDGVKKTLKVVPEGVRVNKTVAVRTLNPE 949
C3_human      KSSLSVPYVIVPLKTGLQEVEVKAAVYHHFISDGVRKSLKVVPEGIRMNKTVAVRTLDPE 950
              :**::****:**** * :********:::******:* *******:*:*****::**:**

              ---------------- CUB ----------------→-------- TED --------
C3_mouse      KLGQGGVQKVDVPAADLSDQVPDTDSETRIILQGSPVVQMAEDAVDGERLKHLIVTPAGC 1010
C3_rat        HLNQGGVQREDVNAADLSDQVPDTDSETRILLQGTPVAQMAEDAVDGERLKHLIVTPSGC 1010
```

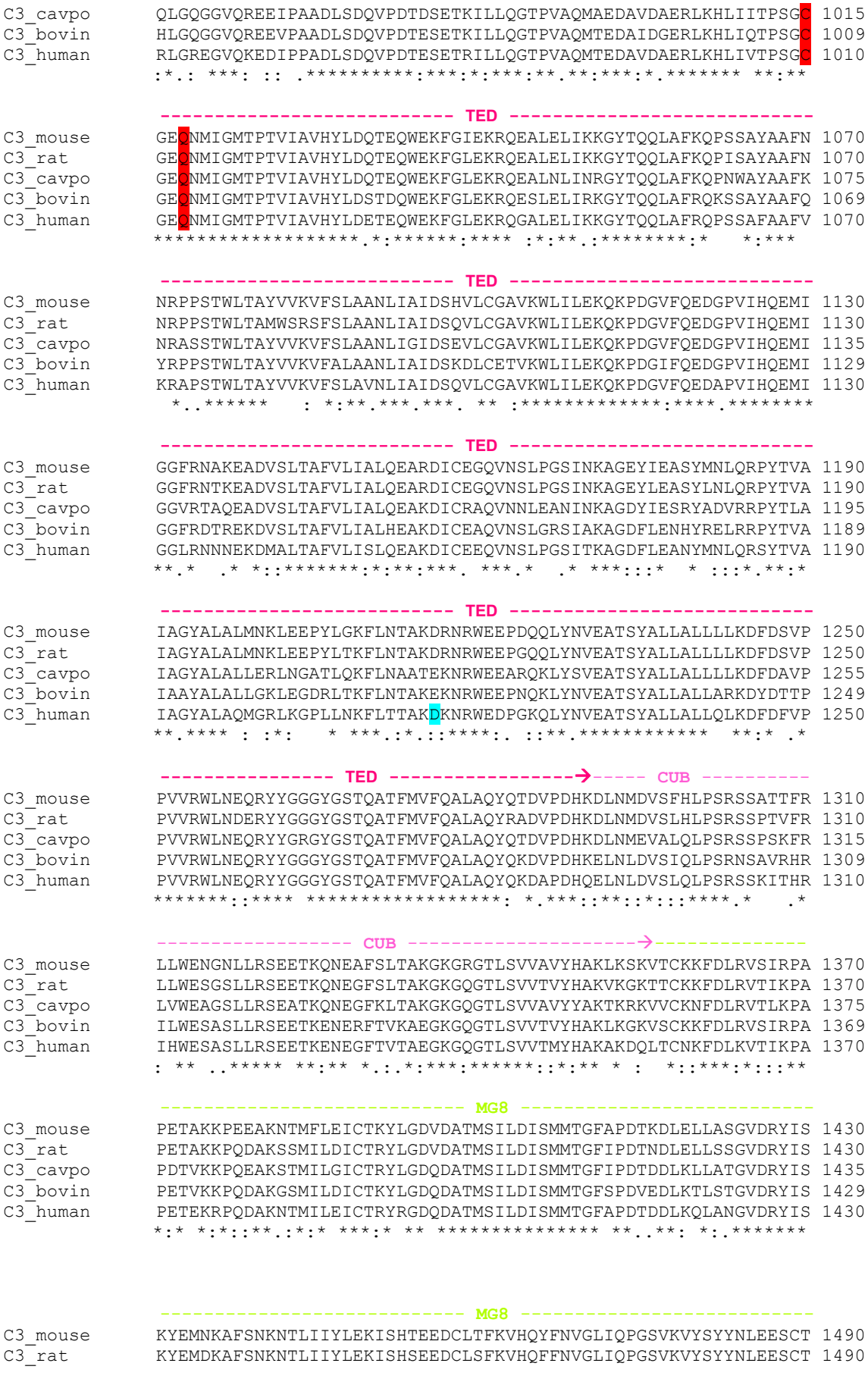

C3_cavpo QLGQGGVQREEIPAADLSDQVPDTDSETKILLQGTPVAQMAEDAVDAERLKHLIITPSGC 1015
C3_bovin HLGQGGVQREEVPAADLSDQVPDTDSETKILLQGTPVAQMTEDAIDGERLKHLIQTPSGC 1009
C3_human RLGREGVQKEDIPPADLSDQVPDTDSETRILLQGTPVAQMTEDAVDAERLKHLIVTPSGC 1010
TED
C3_mouse GEQNMIGMTPTVIAVHYLDQTEQWEKFGIEKRQEALELIKKGYTQQLAFKQPSSAYAAFN 1070
C3_rat GEQNMIGMTPTVIAVHYLDQTEQWEKFGLEKRQEALELIKKGYTQQLAFKQPISAYAAFN 1070
C3_cavpo GEQNMIGMTPTVIAVHYLDQTEQWEKFGLEKRQEALNLINRGYTQQLAFKQPNWAYAAFK 1075
C3_bovin GEQNMIGMTPTVIAVHYLDSTDQWEKFGLEKRQESLELIRKGYTQQLAFRQKSSAYAAFQ 1069
C3_human GEQNMIGMTPTVIAVHYLDETEQWEKFGLEKRQGALELIKKGYTQQLAFRQPSSAFAAFV 1070
TED
C3_mouse NRPPSTWLTAYVVKVFSLAANLIAIDSHVLCGAVKWLILEKQKPDGVFQEDGPVIHQEMI 1130
C3_rat NRPPSTWLTAMWSRSFSLAANLIAIDSQVLCGAVKWLILEKQKPDGVFQEDGPVIHQEMI 1130
C3_cavpo NRASSTWLTAYVVKVFSLAANLIGIDSEVLCGAVKWLILEKQKPDGVFQEDGPVIHQEMI 1135
C3_bovin YRPPSTWLTAYVVKVFALAANLIAIDSKDLCETVKWLILEKQKPDGIFQEDGPVIHQEMI 1129
C3_human KRAPSTWLTAYVVKVFSLAVNLIAIDSQVLCGAVKWLILEKQKPDGVFQEDAPVIHQEMI 1130
TED
C3_mouse GGFRNAKEADVSLTAFVLIALQEARDICEGQVNSLPGSINKAGEYIEASYMNLQRPYTVA 1190
C3_rat GGFRNTKEADVSLTAFVLIALQEARDICEGQVNSLPGSINKAGEYLEASYLNLQRPYTVA 1190
C3_cavpo GGVRTAQEADVSLTAFVLIALQEAKDICRAQVNNLEANINKAGDYIESRYADVRRPYTLA 1195
C3_bovin GGFRDTREKDVSLTAFVLIALHEAKDICEAQVNSLGRSIAKAGDFLENHYRELRRPYTVA 1189
C3_human GGLRNNNEKDMALTAFVLISLQEAKDICEEQVNSLPGSITKAGDFLEANYMNLQRSYTVA 1190
TED
C3_mouse IAGYALALMNKLEEPYLGKFLNTAKDRNRWEEPDQQLYNVEATSYALLALLLLKDFDSVP 1250
C3_rat IAGYALALMNKLEEPYLTKFLNTAKDRNRWEEPGQQLYNVEATSYALLALLLLKDFDSVP 1250
C3_cavpo IAGYALALLERLNGATLQKFLNAATEKNRWEEARQKLYSVEATSYALLALLLLKDFDAVP 1255
C3_bovin IAAYALALLGKLEGDRLTKFLNTAKEKNRWEEPNQKLYNVEATSYALLALLARKDYDTTP 1249
C3_human IAGYALAQMGRLKGPLLNKFLTTAKDKNRWEDPGKQLYNVEATSYALLALLQLKDFDFVP 1250
TED CUB
C3_mouse PVVRWLNEQRYYGGGYGSTQATFMVFQALAQYQTDVPDHKDLNMDVSFHLPSRSSATTFR 1310
C3_rat PVVRWLNDERYYGGGYGSTQATFMVFQALAQYRADVPDHKDLNMDVSLHLPSRSSPTVFR 1310
C3_cavpo PVVRWLNEQRYYGRGYGSTQATFMVFQALAQYQTDVPDHKDLNMEVALQLPSRSSPSKFR 1315
C3_bovin PVVRWLNEQRYYGGGYGSTQATFMVFQALAQYQKDVPDHKELNLDVSIQLPSRNSAVRHR 1309
C3_human PVVRWLNEQRYYGGGYGSTQATFMVFQALAQYQKDAPDHQELNLDVSLQLPSRSSKITHR 1310
CUB
C3_mouse LLWENGNLLRSEETKQNEAFSLTAKGKGRGTLSVVAVYHAKLKSKVTCKKFDLRVSIRPA 1370
C3_rat LLWESGSLLRSEETKQNEGFSLTAKGKGQGTLSVVTVYHAKVKGKTTCKKFDLRVTIKPA 1370
C3_cavpo LVWEAGSLLRSEATKQNEGFKLTAKGKGQGTLSVVAVYYAKTKRKVVCKNFDLRVTLKPA 1375
C3_bovin ILWESASLLRSEETKENERFTVKAEGKGQGTLSVVTVYHAKLKGKVSCKKFDLRVSIRPA 1369
C3_human IHWESASLLRSEETKENEGFTVTAEGKGQGTLSVVTMYHAKAKDQLTCNKFDLKVTIKPA 1370
MG8
C3_mouse PETAKKPEEAKNTMFLEICTKYLGDVDATMSILDISMMTGFAPDTKDLELLASGVDRYIS 1430
C3_rat PETAKKPQDAKSSMILDICTRYLGDVDATMSILDISMMTGFIPDTNDLELLSSGVDRYIS 1430
C3_cavpo PDTVKKPQEAKSTMILGICTRYLGDQDATMSILDISMMTGFIPDTDDLKLLATGVDRYIS 1435
C3_bovin PETVKKPQDAKGSMILDICTKYLGDQDATMSILDISMMTGFSPDVEDLKTLSTGVDRYIS 1429
C3_human PETEKRPQDAKNTMILEICTRYRGDQDATMSILDISMMTGFAPDTDDLKQLANGVDRYIS 1430
MG8
C3_mouse KYEMNKAFSNKNTLIIYLEKISHTEEDCLTFKVHQYFNVGLIQPGSVKVYSYYNLEESCT 1490
C3_rat KYEMDKAFSNKNTLIIYLEKISHSEEDCLSFKVHQFFNVGLIQPGSVKVYSYYNLEESCT 1490

Figure 1.10 The sequence alignment for C3 from five different species. Sequences were aligned using Clustal W (Thompson *et al*., 1994). The signal peptides are included (before MG1). The α-chain includes ANA domain (residue human 665) and onwards. The thioester is formed by the side chain of the highlighted red residues in TED domain (Tack *et al.*, 1980; Thomas *et al.*, 1982). The C3S and C3F variant is highlighted in cyan. The NCBI accession numbers are: C3_mouse P01027; C3_rat P01026; C3_cavpo (guinea pig) P12387; C3_bovin Q2UVX4 and C3_human P01024. Secondary structure elements refer to human C3.

human C3/C4 and C3/α2M have ≈29% and ≈23.5% amino acid identity. The anaphylatoxins C3a and C4a are very similar; all six cysteine residues are conserved. The sites of activation at the COOH termini of C3a and C4a are not conserved in α2M, but the locations of the thioester sites are almost identical. This further emphasises the high degree of similarity among the three proteins. They have evolved from a common ancestor (De Bruijn & Fey, 1985).

The crystal structures of human complement components C3 have been determined (Janssen *et al*., 2005). The structure of C3 displays an intricate arrangement of 13 domains. The core of C3 is formed by eight homologous macroglobulin (MG) domains, together with a thioester-containing domain (TED) domain containing the C3 active site, a C1r/C1s-Uegf-Bmp1 (CUB, protein module initially found in "complement subcomponents C1r/C1s, Uegf and Bone morphogenetic protein-1") domain that links the TED and MG domains, and the C345C, LNK and anaphylatoxin domains. In C3b, the anaphylatoxin domain (C3a) is lost. So far, four linkages of disulphide bridge were determined in the C3a part of C3, and a single bridge was determined in the β-chain as well as in the C3d part of the α-chain (Janssen *et al*., 2005). The N- and C-terminal regions of the α-chain were found to be connected with each other with a disulphide bridge.

The first five domains, MG1-5, are formed by residues of the β-chain, residues 1-534 (Figure 1.11). MG6 is formed by residues from the intertwined β- and α-chain, where residues 535-577 of the β -chain form one half and residues 578 - 642 of the α-chain form the other half. An insertion is located in MG6 and is formed by residues 578-745. This segment includes the linker region (LNK), the tetra-arginine pro-C3 processing site, the anaphylatoxin (ANA) domain and a linker (α'NT) that connects the ANA domain back to MG6. The last two MG domains, MG 7-8, on the α–chain, are separated by the CUB domain and the thioester-bond containing (TED) domain, and the carboxy-terminal 170 residues form the 'anchor region' and the C345C domain.

The eight MG domains without sequence homology exhibit a fibronectin-type-3-like core fold (Figure 1.12). MG8 corresponds to the receptor-binding domain in α2M; as expected, the structure of MG8 is homologous to this domain. The MG1-6 domains form 1.5 turns of helical coil as in a key ring (Figure 1.13). MG7 and MG8 on the α-chain are outside

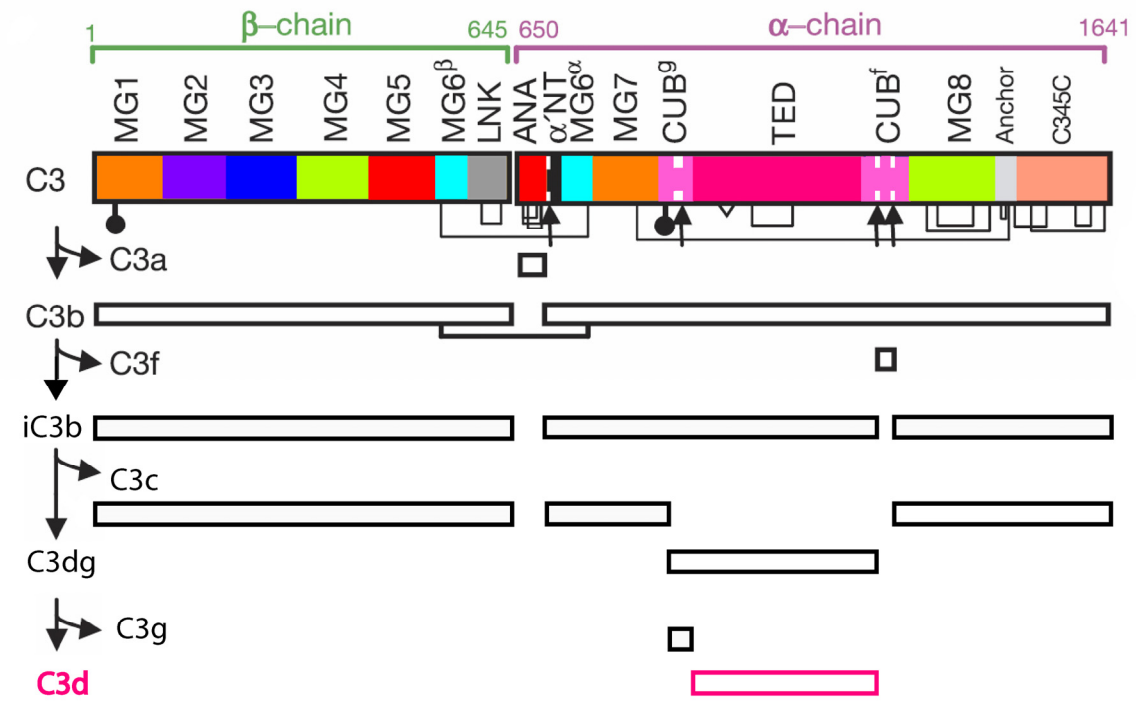


Figure 1.11 Domain sequence and arrangements in C3. Thioester site shows as white triangle, disulphide bridges are connected (adapted from Janssen *et al*., 2005). The first residue of β-chain is labelled as "1". The colours matches Figure 1.10.

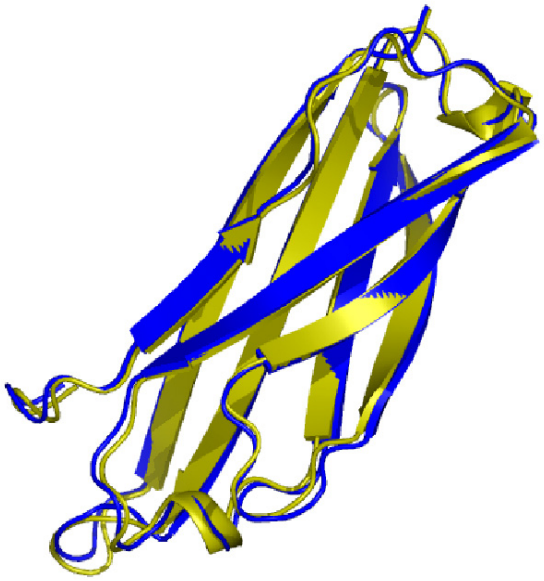


Figure 1.12 An *MG* domain (Janssen *et al*., 2005).

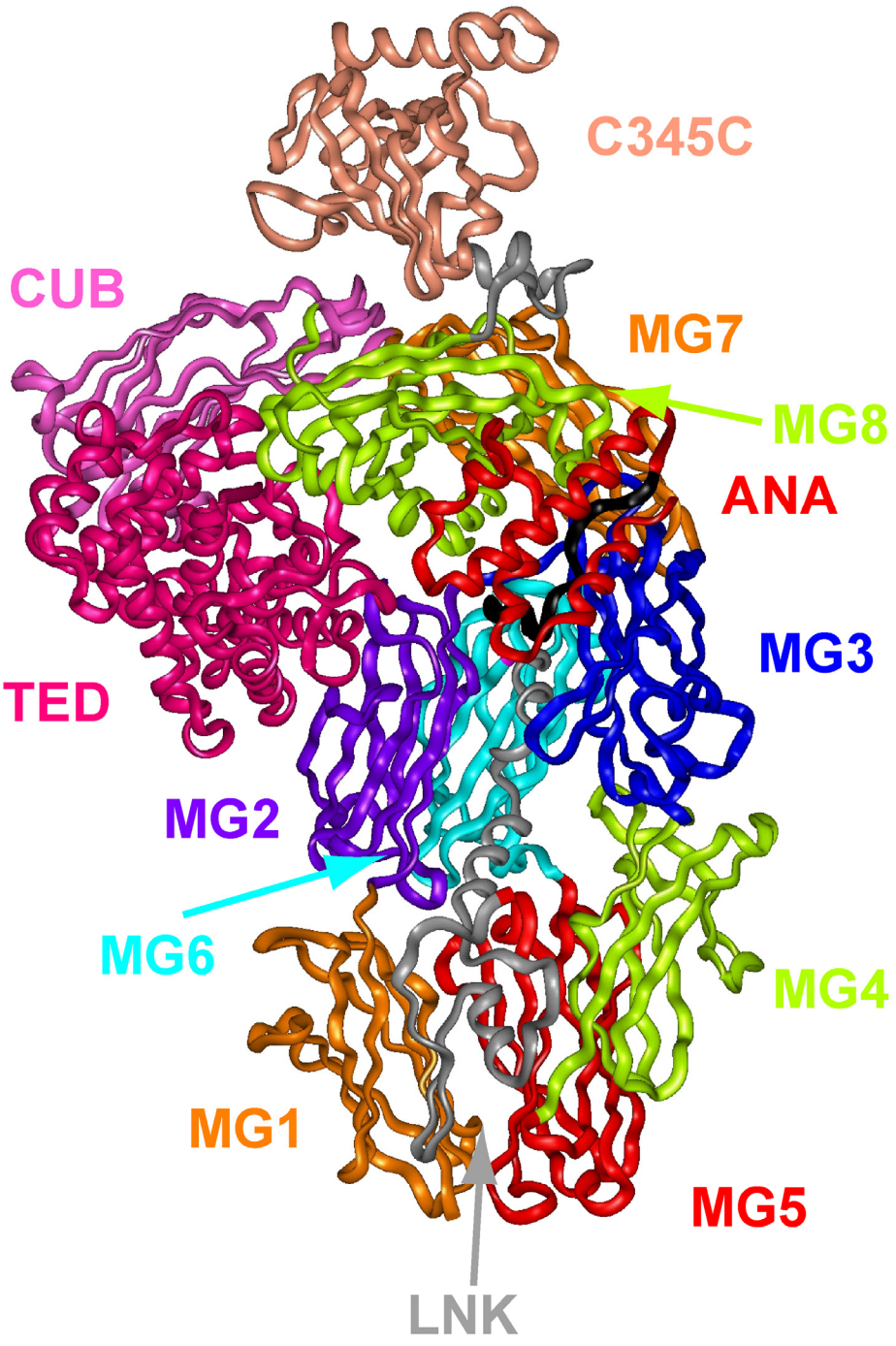


Figure 1.13 The crystal structure of C3. INSIGHT II 98.0 molecular graphics software (Accelrys, San Diego, CA, USA) was used to create the ribbon structure from PDB entry 2a73 (Janssen *et al*., 2005).

of the key ring structure. They are close to each other and mismatch slightly in C3.

LNK, ANA and α'NT domains are the insertions between $MG6^{\beta}$ and $MG6^{\alpha}$. The LNK domain is wedged in between domains MG1, MG4 and MG5. It has three aromatic residues from the third helix and a β-strand forms a small hydrophobic core. The ANA domain when cleaved off forms the C3a fragment. The structure of this domain in C3 is similar to the crystal structure of C3a. Following this, there is a scissile bond Arg 726- Ser 727 in a disordered loop and an extended loop (residue 730-745) that connects ANA to MG6. Loop 727-745 is referred as α'NT. The geometric arrangements of these insertions are very interesting. The insert loops completely through the central hole formed by the MG1-6 domains. C3 appears as two chains that are intertwined in domain MG6. The C-terminus of the β-chain and N-terminus of the α-chain are far apart; there is a gap of 5.3 nm between the last β residue (Pro-643) and first α residue (Val-651).

Residues 913-1334 form a segment that inserts between domains MG7 and MG8. These residues form the CUB domain and the TED domain. The TED domain itself (residues 963-1268) is inserted in loop of the CUB domain. Consequently, CUB is an intertwined domain consisting of two distinct parts of the polypeptide chain (residues 912-962 and 1269-1330). The cleavage fragments C3d, C3g and C3f are part of this CUB-TED insert.

At the C-terminal end the C345C domain foms a knob-like protrusion (Figure 1.13). This is connected to MG8 via a short anchor region. The feature is shared in Complement proteins C3, C4 and C5, and not known in any other members of the α2M family. The anchor region links the C345C domain both to the MG7 domain, by a disulphide bridge, and to the MG8 domain by the polypeptide chain.

(1.5.3) Function of C3 in complement

Human C3 in the humoral immune system plays at least three different biological functions. Firstly, the activation of C3 leads to target cell lysis through the complement activation pathways. The initiation of C3 cleavage leads to covalent C3b deposits onto target cell surfaces, thus leading to the membrane-attack complex on target cells. Secondly, human C3 is involved in antigen presentation and consequently facilitates cellular immune

response. It has been shown that cross-linking of C3b to tetanus toxin reduces the amount of antigen required to induce specific T cell responses (Rey-Millet *et al*., 1994). Thirdly, C3 functions in cell regulation. During C3 activation, different C3 fragments are generated that react with specific cell-surface receptors, such as C3aR, CR1, CR3, CR4 and CR2, consequently to trigger regulation of cellular functions. As a result, normal cells or tumour target cells against autologous complement lysis by preventing deposition of activated C3b on their surface.

(1.5.4) C3 mutations and disease

Since 1972, inherited C3 deficiency has been described in a variety of ethnic and national origins (Singer *et al*., 1994a, 1994b). Most patients show a heightened susceptibility to bacterial infections, thus leading to purulent lesions. C3 deficiency normally leads to impaired chemotactic activity, slow response of neutrophils to infectious agents and a series of immune complex diseases, such as systemic lupus erythematosus and membrane proliferative glomerulonephritis. In patients, defective antibody responses were observed, which suggests C3 may be critical in the generation of normal immune response (Singer *et al*., 1994a, 1994b; Bitter-Suermann & Burger, 1990).

Recently, the molecular genetic basis of C3 deficiency has been examined. There are at least six types of mutations. Two of the patients show point mutations at 5'-donor splice sites in the C3 gene. An English male has a site mutation in intron 18 of C3 gene. The mutation generates a 61-nucleotide deletion in the C3 RNA during RNA processing. The deletion causes a shift in reading frame of the C3 transcript and a premature stop in translation (Botto *et al*., 1990). The other site mutation described in a Taiwanese aboriginal female (Huang & Lin, 1994): a 5'-donor splice site mutation in intron 10 causes exon 10 to be skipped during RNA processing, consequently a premature stop codon was generated. A third case is from the Afrikaaner population of South Africa (Botto *et al*., 1992), where an inherited C3 deficiency found in a family carries an 800 bp deletion in the C3 gene (includes exons 22 and 23). A fourth case is a missense mutation in exon 13 that results in a critical amino acid substation in C3 β-chain (Asp^{549} to Asn). This Asn^{549} substitution causes impaired secretion of the intracellular C3 proteins (Singer *et al*., 1994a, 1994b). A C3 deficiency also was found in a Laotian family and results from reduced levels of C3-specific mRNA (Singer *et al*., 1996), although the exact molecular genetic defect for the

fifth case is yet to be clarified. The sixth case is reported recently in relation to atypical haemolytic uremic syndrome (aHUS) (Frémeaux-Bacchi *et al*., 2008). Nine mutations were found in aHUS patients with low serum C3 level. There were eight heterozygous missense mutations, R570W, R570Q, R713W, A1072V, D1093N, C1136W, Q1139K, and H1142D and one heterozygous nonsense mutation Y832X. By cofactor activity assay, five of these mutations have decreased binding affinity with MCP and two have normal binding affnity, therefore C3 could be a susceptibility factor for aHUS.

(1.5.5) C3 fragments

In three complement pathways, the enzyme complex C3-convertase activates C3 by the proteolytic cleavage of C3 into C3a (= ANA) and C3b (remaining 12 domains). C3a is a vasoactive peptide and a mediator of inflammation (Hugli, 1975, 1981). The removal of C3a induces a conformational change in the C3b portion of molecule. The α'NT domain shifts from one face of the protein to the other; as a consequence, the MG8, CUB and TED domain undergo a large relocation in relation to the protein core MG1-6 (Janssen *et al*., 2006). Furthermore, relocation leads to the exposure of an internal thioester bond, which is buried and quite inaccessible in native C3. By this thioester bond, C3b could covalently bind to the cell by an ester or amide bond and thus ultimately lead to formation of the membrane-attack complex on target cells, thus inducing inflammation and elimination of self and non-self targets.

Without any enzyme present, C3 is also continuously activated at a slow rate in the fluid phase to become C3u (C3(H_2O)). C3u can be activated by enzyme complexes which cleave C3 into anaphylatoxin C3a and the major fragment C3b. It has structural and a functional similarity with C3b. C3a consists of the first 77 amino acids of the α-chain of C3 and mediates inflammation (Bokisch *et al*., 1969; Figure 1.11). C3b is a 179 kDa protein and can covalently attach to antigen surfaces as a component of C3 and C5 convertases, and modulate the activities of the C3b as an opsonin for cellular receptors. Subsequent proteolysis of C3b results in the formation of the fragments iC3b and C3f, and finally the two fragments C3dg and C3c (Figure 1.11). iC3b and C3dg may stimulate B-cells providing a link between the innate and adaptive immune response (Caroll, 2004). C3f is a weak spasmogen that functions like C3a anaphylotoxin (Ganu *et al*., 1989). C3d contains the thioester-forming residues, and is equivalent to the TED domain (residues 1205-1214) of C3

(Figure 1. 11). C3d is less accessible in C3 and C3b, so that these molecules have low affinity for CR2. C3d can be covalently attached to an antigen, and has the ability to significantly lower the amount of antigen required to elicit an antibody response (Dempsy *et al*., 1996; Green *et al*., 2003).

(1.5.6) Structure of C3b and C3u

So far, five C3b crystal structures are available (Janssen *et al*., 2006; Wiesmann *et al*., 2006; Wu *et al*., 2009; Rooijakkers *et al*., 2009; Ajees *et al*., 2006). In C3b, four crystal structures show that the TED and CUB domains are extended along the major axis of the MG domains and in contact with these MG domains (Janssen *et al*., 2006; Wiesmann *et al*., 2006; Wu *et al*., 2009; Rooijakkers *et al*., 2009), while a fifth deviant crystal structure shows that the TED and CUB domains are significantly extended away from the MG domains (Ajees *et al*., 2006). Even though the crystallographic evidence for the fifth C3b structure was disputed, the criticisms of this structure were not accepted (Janssen *et al*., 2007; Ajees *et al*., 2007). Most recently, the Ajees *et al.* crystal structure was reported to be falsified (Borrell, 2009). Electron microscopy studies of C3b *in vacuo* also supported the location of the TED and CUB domains seen in the four C3b crystal structures (Nishida *et al*., 2006). H-D exchange rates that were measured in mass spectrometry studies of C3 and C3b indicated that a large conformational change occurs between these two proteins, while C3u and C3b show similarities with each other except the additional anaphylatoxin domain (Schuster *et al*., 2008).

An unresolved question in relation to C3b is whether there is any basis for the deviant C3b structure. Another involves the structural transition between C3 and C3b. Recent developments in X-ray scattering have resulted in considerably improved signal-noise ratio at the ESRF facility, alongside with constrained modelling methods that result in structural determinations that are deposited in the Protein Data Bank (Bonner *et al*., 2009; Perkins *et al*., 2009). In Chapter 5, I combine analytical ultracentrifugation and X-ray scattering to show that conformational and self-association differences exist between C3 and C3u, and apply constrained modelling to determine their best-fit solution structures, in order to reveal new insights into C3 activation.

(1.6) Factor H

In the alternative pathway, Factor H (FH) has cofactor activity converting C3b to iC3b. It is found at the RCA gene cluster (same as CR2) and encoded by a single gene (*HF1*) (Rodriguez de Córdoba *et al*., 1999). It is a 155 kDa glycoprotein (Fenaille *et al*., 2007) that circulates in human plasma at a concentration of 235 – 810 μg/ml (Saunders *et al*., 2006). The polymorphism Tyr402His of FH is a major risk factor for developing age-related macular degeneration (AMD). AMD is the most common cause of visual impairment in the elderly in Western countries.

FH is mainly produced by the liver (Schwaeble *et al*., 1987) as a single chain molecule of 1231 amino acid residues (Ripoche *et al*., 1988). The first 18 amino acid is the signal peptide; after secretion, the mature human FH contains 1213 amino acid residues. Like CR2, FH contains multiple SCR domains, with only β-strands and no α-helix in the secondary structure (Morikis & Lambris, 2005). There are 20 SCR domains with 19 flexible linkers (3-8 residues) between domains (Figure 1.14a). SCR domains have conserved four Cys and a Trp, with an exception at SCR 10. Two disulphide bonds are present in the pattern of Cys^1-Cys^3 and Cys^2-Cys^4. There are also glycine, proline and tyrosine conserved in the SCR sequences (Aslam & Perkins, 2001; Morikis & Lambris, 2005; Saunders, *et al.,* 2006). FH is heavily glycosylated including Asn217 in SCR-4, Asn529 in SCR-9, Asn718 in SCR-12, Asn802 in SCR-13, Asn822 in SCR-14, Asn882 and Asn911 in SCR-15, Asn1029 in SCR-17, and Asn1095 in SCR-18 (Figure 1.14). Except the site at Asn217, the other eight sites are proved to be glycosylated by diantennary disialylated glycans of 2204 Da by mass spectrometry (Fenaille *et al*., 2007).

(1.6.1) Self-association of Factor H

FH oligomerisation in solution was first identified in 1991(Perkins *et al*., 1991). So far several FH fragments have been reported to self-associate. Dimers of fragment SCR 15-18 and SCR 15-20 have been seen in non-reducing sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and Western blotting analysis (Jokiranta *et al*., 2000). An interaction between the FH N-terminal fragment SCR 1-7 and intact FH was observed by SPR (Oppermann *et al*., 2006). Fragment SCR 6-8 and 16-20 in solution has been found to self-associate in AUC analyses (Fernando *et al*., 2007; Okemefuna *et al*., 2008). Polyanion

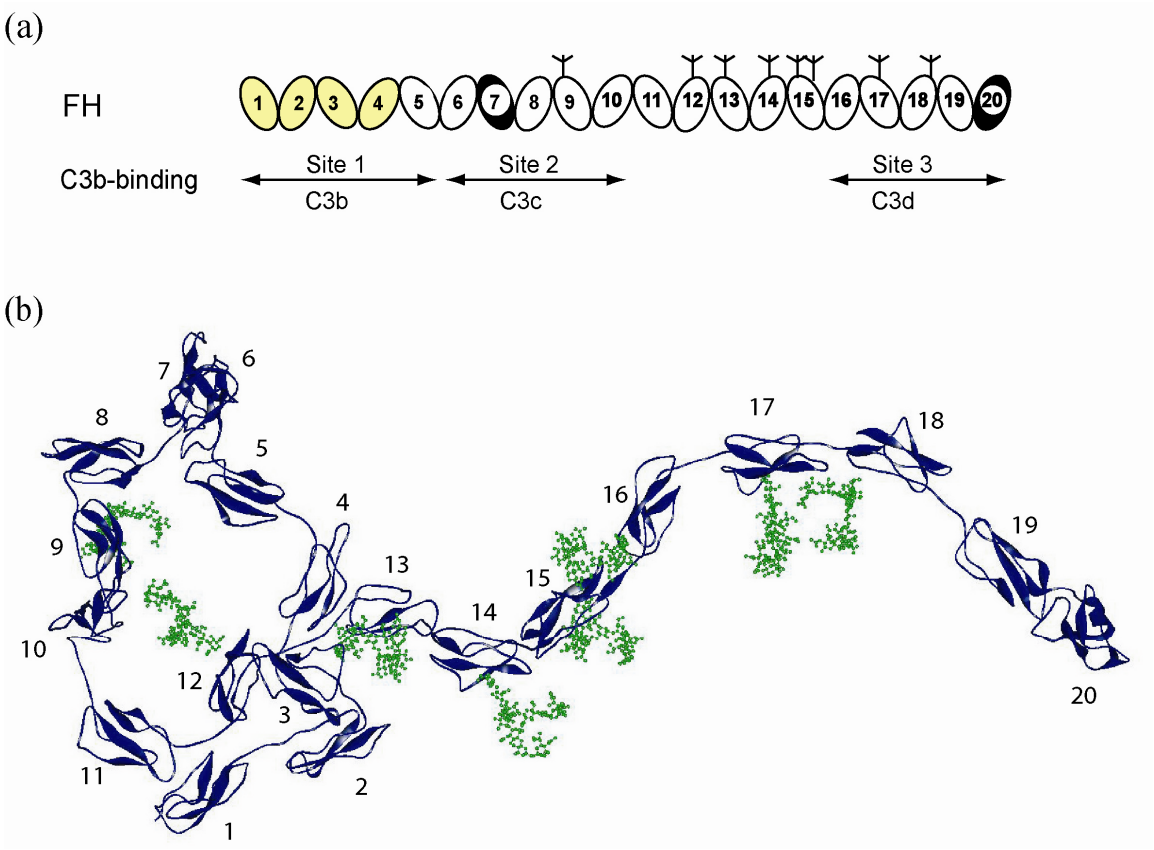


Figure 1.14 A schematic diagram of FH binding sites (adapted from Okemefuna, 2009a). (a) Schematic view of the 20 SCR domains of FH with different ligands binding sites for some C3 fragments. The first region also has decay acceleration and cofactor activity (yellow coloured domains). Two heparin-binding sites on SCR-7 and SCR-20 (black), and eight putative N-linked glycosylation sites (Y) are shown. (b) FH has flexible beads-on-string domain arrangement (PDB code 3GAV, Okemefuna *et al*., 2009a).

such as dextran sulphate or heparin was suggested to induce FH to form tetramers on SCR 18-20 (Pangburn *et al*., 2009).

The most recent study showed that a reversible monomer-dimer equilibrium of intact FH exists up to 1.3 mg/ml, and then more oligomers appear with increased concentration (Nan *et al*., 2008b). Similar studies were carried out in a range of NaCl concentrations and pH values. The proportions of FH monomer, dimer and higher oligomers in 137 mM NaCl were about 84%, 8% and 8% (Okemefuna *et al*., 2009a). Both studies showed dissociation constant K_D value of 16 - 28 μM at pH 7.4, 137 mM NaCl with dimer 5% to 23% in blood. The FH oligomer formation study indicates that the regulatory role of FH will be affected by the local environment and its effect on FH oligomers.

(1.6.2) C3b binding to Factor H

The partially folded-back arrangement of 20 SCR domains in FH also showed a dependence on electrostatic effects (Okemefuna *et al*., 2009a). This may be important for the mechanism of its regulatory action through interaction with C3b. The complement amplification step is initiated by the formation of the C3 convertase through the binding of factor B to C3b and subsequent proteolytic activation of factor B. FH is the regulation the alternative pathway of C3 convertase C3bBb and its C3b component, both in plasma and on host cell surface (Pangburn *et al*., 1977; Whaley & Ruddy, 1976). FH and factor I prevent convertase formation and trigger the proteolytic inactivation of C3b. FH regulates surface-bound C3b activity by recognising charge clusters on the surfaces that are mimicked by heparin, then it makes the initial contact with target through the C-terminal SCR-20 (Oppermann *et al*., 2006; Ferreira *et al*., 2006).

There are three C3b binding sites on FH (Figure 1.14a). C3b binds with FH N-terminal domain SCRs 1-4, where the cofactor and decay accelerating activities of FH are located. The second site is on FH SCR-6/10, where it binds to the C3c region of C3b. A third site within SCR-16/20 binds to the C3d region of C3b (Lambris *et al*., 1988; Sharma & Pangburn, 1996; Jokiranta *et al*., 2000; 2001). The known heparin binding sites are SCR 7, SCR 13 and SCR 20 (Blackmore *et al*., 1998; Pangburn *et al*., 1991). It is possible that these three binding sites for C3b are not independent of each other but work to bind C3b in a synergistic manner. The stoichimetry of C3b and FH may not be a simple 1:1 complex.

In the alternative pathway, SCR 1-4 may compete with fragment Bb to bind to C3b on the same region (Wu *et al*., 2009). A previous study showed that the α'NT and MG7 regions may be a primary binding site for FH. (Wu *et al*., 2009). The second main binding site may involve a conserved hydrophobic residues surrounded by hydrophilic residues on SCR2 that interact with MG6 of C3b. Finally, the hypervariable loop of SCR3 of FH may interact with the CUB domain of C3b, because on the crystal structure of SCR 1-4 with C3b, the SCR2-SCR3 part of FH bound to C3b, with the hypervariable loop of SCR3 directly contacting the CUB domain. A molecular understanding of the FH-C3b complex requires the dissection of the role of each of these C3b sites on FH.

C3d is similar to the TED part of C3 with a negative surface charge. Surface plasmon resonance analysis in 50 mM and 75 mM NaCl indicated that C3d binds weakly to FH and the FH-related proteins FHR-3 and FH-4 with dissociation constants K_D of 0.09-0.36 μM (Hellwage *et al*., 2002). FH-C3d multimers with possible stoichimetries of 1:1, 2:1 or 2:2 have been seen in both the solution and surface phases using physiologically-relevant concentrations and different salt concentration buffers (Okemefuna *et al*., 2008). Because FH binds with C3d part of C3b on SCR 19/20, the knowledge of C3d-FH indicates that the C3b-FH interaction is comparatively weak.

From surface plasma resonance, we knew that the interaction between C3b and FH is relatively weak (Schmidt *et al*., 2008). Studies of the complex of C3d-FH will help us understand the possibility of how C3b binds with FH in solution. In this thesis, insight into C3b interaction with FH was obtained by studying the interaction between C3 and C3u in different salt concentration (Chapter 5) and likewise with FH present to alter the strength of the C3u-FH interaction in solution (Chapter 6). The effect of ionic strength on these interactions provides new insight into complement regulation and host protection.

Chapter Two

Analytical ultracentrifugation

(2.1) Introduction to protein structure determination techniques

Over the past few decades, much research efforts has been dedicated to protein structure determination. It was realised that the structural information for proteins contained a wealth of biomedical information, which among other things assists in the development of new drugs for all kinds of diseases. Several techniques have been developed to determine protein structures.

Protein amino acid primary structure can be determined directly by mass spectrometry, or Edman degradation, Sanger's reagent or Dansyl chloride methods. It can also be determined indirectly through sequencing the protein genes or cDNA. The secondary structure of a protein can be determined by circular dichroism (CD), Fourier transform infrared (FT-IR) spectroscopy and nuclear magnetic resonance (NMR) spectroscopy (Latham *et al*., 2005). For protein tertiary structure, there are techniques with low or high structural resolution. The low resolution techniques (details of 2-4 nm) include analytical ultracentrifugation (AUC), small angle X-ray and neutron scattering (SAXS and SANS; Chapter 3) and electron microscopy (EM). The high resolution techniques include nuclear magnetic resonance (NMR) and X-ray crystallography. Overall, each structural determination technique has its advantages and disadvantages. Usually the one which can provide the highest resolution structural information is applied, depending on the system of interest.

AUC, SAXS and SANS are popular methods to study macromolecules in solution. AUC has advantages for studying the sizes and shapes of macromolecules, in particular the state of oligomerisation. SAXS and SANS can study macromolecules in solutions by the diffraction of X-ray or neutrons respectively. The structural resolution of AUC, SAXS and SANS can be improved to medium levels (0.5-1.0 nm) if combined with biomolecular modelling methods.

EM is a low resolution technique that the operating principles is similar to light microscopy, except it uses a beam of electrons and magnetic lense to visualise the conformation of protein molecules. The greater resolution of EM compared to light microscopy is because of its usage of electron wavelengths about 100,000 times shorter than visible light. However, the high level of electron radiation may damage the protein

sample. In addition, samples of hydrated materials have to be prepared in various ways in order to stabilise them, and to increase their electron optical contrast (staining) to permit visualisation *in vacuo*. These processes may result in damage to distort the protein structures, particularly their structural arrangement in physiological condition. Hence structural information obtained under such harsh conditions may not fully represent the protein's native conformation. Furthermore, low signal-to-noise ratio means that structural information from several images have to be combined and averaged, thus this creates difficulties in presenting highly flexible structures. This is particularly true for flexible multiple domain molecules with variable interdomain linkers such FH, CR2 and CR1.

NMR is a field of spectroscopy based on the absorption of electromagnetic radiation in the radio-frequency region. It provides information on the topology, dynamic and three-dimensional structure of molecules in solution. In biology, NMR uses the magnetic properties of some nuclei, mostly ^{1}H, ^{15}N and ^{13}C. When a strong magnetic field is applied to a protein solution, the net spin aligns in the direction of the field. Then a second magnetic field is applied to the sample, and perpendicular to the direction of the first field. The nuclei are excited. When the nuclei return to their original alignment, radiation is emitted. This dependence of nuclear magnetic energy levels on the local electronic environment is known as the chemical shift. NMR experiments apply magnetic fields at different radio frequencies to labelled or non-labelled protein samples, and monitor their chemical shifts. Currently, multinuclear NMR permits the assignment of chemical shifts to specific nuclei, and Nuclear Overhauser effects are used to determine spatial restraints between atoms. Informations from them allow three dimensional models to be created for the protein structure (Wüthrich, 1986). NMR studies require protein molecules equivalent or smaller than 35 kDa, therefore it is an inappropriate technique for studying the intact form of relatively large proteins such as FH or CR2.

X-ray crystallography is a technique of determining the arrangement of atoms by interpreting the diffraction of X-rays from the repeating unit cells that constitute the crystal lattice. Protein crystals are fragile and maintained by weak hydrogen bond and ionic forces. Proteins with flexible conformations and many glycosylation sites are sometime hard to crystallise. It is difficult to grow intact CR2 and FH crystals for both reasons. Provided the protein crystals can be grown, this method can yield high resolution structures (0.1-0.3 nm).

Crystals are exposed to a monochromatic, collimated X-ray beam, diffraction data are acquired, and electron density maps are calculated, to which the structure is fitted.

(2.2) Introduction on analytical ultracentrifugation

AUC is a powerful tool to characterise the macromolecules in their native state under biological relevant solution conditions (see examples in Cole & Hansen, 1999). An analytical ultracentrifuge combines the functions of an optical detection system and a preparative ultracentrifuge. The macromolecular samples are prepared in specially designed sample cells and are put into an ultracentrifuge rotor and subjected to a high centrifugal force. At the same time, spectroscopic methods, which are based on either absorption or refractive index differences between sample and reference buffer in the cell, are applied to track the effect of the centrifugal force. AUC monitors sedimentation events in real time. By applying the appropriate rotor speed, two types of experiments are performed: sedimentation velocity and sedimentation equilibrium. The data obtained are then analysed to identify the oligomeric state of the macromolecule and its polydispersity, and model the shape of macromolecules.

(2.3) Historical review on analytical ultracentrifugation

The use of ultracentrifugation to determine particle structures was first proposed in 1913 by Dumansky (Serdyuk *et al*., 2007). Then Svedberg and Nichols constructed the first centrifuge with an optical system to follow particle behaviour in a centrifugal field (Svedberg & Nichols, 1923). A year later, Svedberg noted the decrease in absorbance at the top of the cell during centrifugation. Between 1926 and 1927, he determined the molecular weight of haemoglobin using a combination of sedimentation and diffusion data. The behaviour of the moving boundary in the ultracentrifuge was explained by the Lamm equation (Lamm, 1929). The equation is an infinite series of integrals, which can be computed only by numerical integration. In the 1930s, Schlieren optical systems were designed, which allowed a representation of the concentration gradient as a function of distance in the centrifuge sample cell. Physicists started to use Perrin's hydrodynamics theories for ellipsoids of revolution to interpret the frictional coefficient in terms of the shape and hydration of macromolecules. In this period, important achievements in the field of molecular structure included the prediction of the dimensions and shape of tobacco mosaic virus, before the rod-like particles were visualised by electron microscopy.

AUC became commercially available from the 1940s. With the increased use of sedimentation, many achievements in molecular biology were accomplished with the centrifuge. In 1958, Meselson and Stahl use density gradient ultracentrifugation and isotope labelling, to prove the semiconservative mechanism of DNA replication. In late 1959, the ribosome was discovered (Tissières & Watson, 1958; Schachman, 1959). Following the development of the first scanning photoelectric absorption optical system, the ultracentrifuge was able to study proteins, ribosomes, DNA and viruses. Williams, van Holde, Schachman, Yphantis and Fujita developed the essential theory and practice of AUC analysis at this time. In 1970s, the incorporation of a monochromator in the absorption optics allowed extremely dilute solutions to be studied. Rayleigh interference optics yielded highly accurate data for non-absorbing solutes, and was used mainly for the determination of macromolecular mass by sedimentation equilibrium. In 1978, the development of new data analysis methods, which removed the contribution of diffusion from sedimentation velocity boundaries, yielded integral distributions of sedimentation coefficients (Van Holde & Weischet, 1978). Data analysis was much improved with the appearance of a new generation of instruments, which were highly automated for data collection and analysis in the 1990s. The Beckman XL-I ultracentrifuge includes two different optical detection systems: a UV absorption system that makes it possible to study proteins or nucleic acids at very low concentration and a Rayleigh interference optical system, which can detect macromolecules with low or no light absorbance (McRorie & Voelker, 1993). Recently, AUC can be applied to molecules with molecular weights from several hundreds to tens of millions. It can provide accurate determination of the purity, mass, shape, self-association and other binding properties of macromolecules in solution. It is now an indispensable technique in biological laboratories.

(2.4) Theory of sedimentation

A particle suspended in a solvent in an analytical ultracentrifuge is subjected to three forces (Figure 2.1): the sedimentating, buoyant and frictional force. First, there is a sedimenting, or gravitational forces, F_s, proportional to the mass of the particle and the acceleration.

(Eq. 2.1) $$F_s = \omega^2 rm = \omega^2 r \frac{Mr}{N}$$

where m is the mass in grams of a single particle, *Mr* is the molar mass of the solute in g/mol and *N* is Avogadro's number.

Secondly, there is a buoyant force, $F_{b,}$ which is equal to the weight of fluid displaced:

(Eq. 2.2) $F_b = -m_0\omega^2 r$

where m_0 is the mass of fluid displaced by the particle:

(Eq. 2.3) $m_0 = m\bar{v}\rho = \frac{Mr}{N}\bar{v}\rho$

Here, $\bar{v}$ is the volume in m that each gram of the solute occupies in solution (the partial specific volume; the inverse of its effective density) and ρ. is the density of the solvent (g/ml).

Provided that the density of the particle is greater than that of the solvent, the particle will begin to sediment in a radial direction when gravitational force is applied. As the particle begins to move away from the meniscus, its velocity, *u*, will an increase because of the acceleration ($a = \omega^2 r$, r is increasing). As a result, the particle will experience an increased frictional force, $F_{f,}$ (Figure 2.1) acting in an opposite direction to the sedimentation force:

(Eq. 2.4) $F_f = -fu$

f is the frictional coefficient that is proportional to the viscosity of the solvent and the linear dimensions of the particle. An elongated molecule has a greater frictional force than a small, compact one of the same mass. The frictional, buoyant and sedimenting forces reach equilibrium within a short time (less than 10^{-6} s). So the frictional force is equivalent to the difference between the sedimenting and buoyant forces (Serdyuk *et al*., 2007):

(Eq. 2.5) $F_f = r\omega^2\frac{Mr}{N} - \frac{Mr}{N}\bar{v}\rho\omega^2 r$

So

(Eq. 2.6) $fu = r\omega^2\frac{Mr}{N} - \frac{Mr}{N}\bar{v}\rho\omega^2 r$

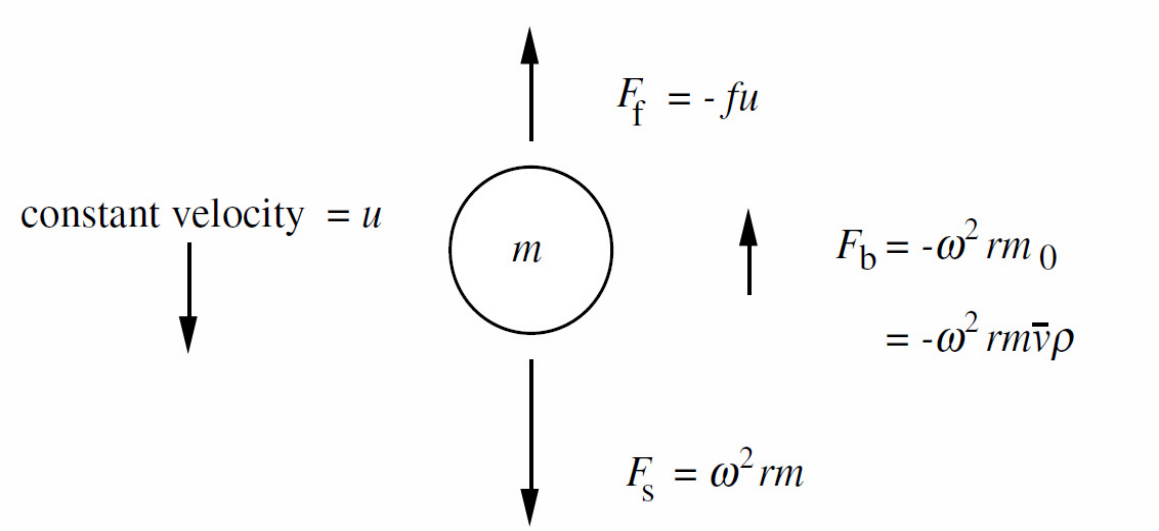


Figure 2.1 The forces acting on a solute particle during centrifugation. F_f is the frictional force. F_b is the buoyant force. F_s is the sedimentating force. F_f and F_b act in an upward direction, while F_s and u act downwards (Ralston, 1993).

Rearranging,

(Eq. 2.7) $$fu = \frac{Mr}{N}\omega^2 r\,(1 - \bar{\upsilon}\rho)$$

Dividing both side of Equation 2.7 by $f\omega^2 r$

(Eq. 2.8) $$\frac{u}{\omega^2 r} = \frac{Mr}{Nf}(1 - \bar{\upsilon}\rho) = s$$

Where $u/\omega^2 r$, the velocity of the particle per unit gravitational acceleration, is called the sedimentation coefficient. The sedimentation coefficient (s) is directly proportional to the buoyant effective molar mass of the particle, and inversely proportional to the frictional coefficient (f). It has a unit of seconds (s), but is conventionally expressed in terms of Svedberg units (S), in recognition of Svedberg's fundamental pioneering work in AUC (Cole & Hansen, 1999). Most sedimentation coefficient values lie between 1 x 10^{-13} s and 100 x 10^{-13} s, and one Svedberg unit is equivalent to 10^{-13} s. Therefore a value of 3 S would translate to 3 x10^{-13} s. Equation 2.8 can be alternatively expressed in standard form as the Svedberg equation:

(Eq. 2.9) $$\frac{Mr}{RT}(1 - \bar{\upsilon}\rho) = \frac{s}{D}$$

R is the gas constant and T is the absolute temperature in Kelvin. D is the diffusion coefficient. In practice, experimentally-observed s values are often converted to standard conditions (at 20°C) by correcting for the buffer, temperature conditions and dilution to zero concentration, and this is presented as $s^0_{20,w}$ (see for example Balbo & Schuck, 2005).

(2.5) Instruments

The most popular analytical ultracentrifuges are the Optima™ or Proteomelab™ XL-A and XL-I systems, both made by Beckman Coulter Instruments. The XL-A uses UV and visible absorbance optics for absorbance measurements, while the XL-I uses integrated absorbance and interference optics for both absorbance and interference measurements (McRorie & Voekler, 1993). More recently, fluorescence optics have been integrated into the XL-I optical system to allow labelled protein in more concentrated solutions to be analysed, such as cell lysates, serum and cell culture media (Cole *et al*., 2008).

The instrument used for my thesis is the Beckman Coulter XL-I analytical

ultracentrifuge. It consists of a rotor system (Section 2.5.1) and an optical detection system (Section 2.5.2). The centrifugation parameters, such as temperature, rotor speed and data acquisition are under computer control.

(2.5.1) The rotor and cells

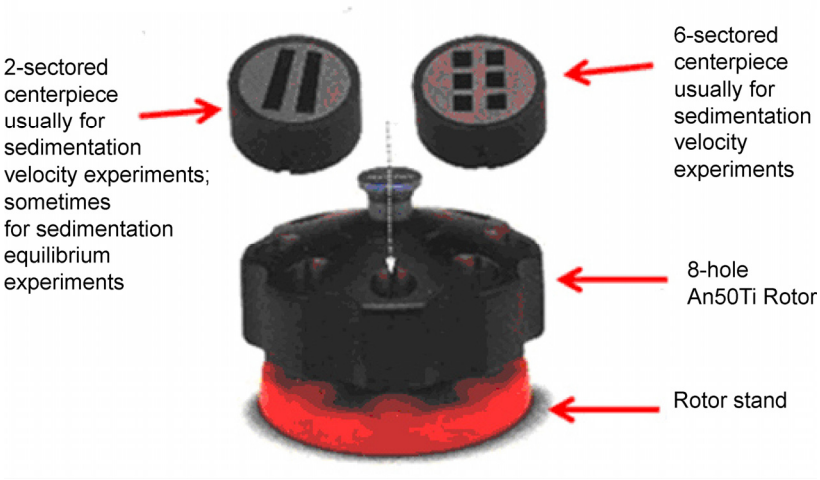
A Beckman Coulter XL-I analytical ultracentrifuge uses specialised rotor systems. To view the sample, the rotor has holes to hold the sample cells (Figure 2.2a). At high speeds, an ultracentrifuge rotor generates a gravitation force in the cell. With a speed of 60,000 r.p.m, a mass of 1g would weight 250,000 times of its weight, 250 kg. Therefore, AUC rotors are capable of withstanding great gravitational forces, and the ultracentrifuge contains a reinforced cylinder chamber in case of rotor breakage. Rotors contains either four or eight holes and are capable of holding either three or seven cells together with a reference cell used for radial calibration purposes. The multiple cells make possible study of up to seven samples in a single experiment.

The centrifuge cells house the sample and reference buffer, and are assembled before each experiment from a cell housing, upper and lower plane windows of optical grade sapphire or quartz, and a centrepiece (Figure 2.2b). In a sedimentation velocity experiment, a double-sector centrepiece is used; while in a sedimentation equilibrium experiment, a six-sector centrepiece is used. Centrepieces are made out of a variety of tough, inert materials, such as epoxy, anodised aluminium or titanium. Aluminium-filled or charcoal-filled epoxies are used for biological experiments at speeds up to 42,000 r.p.m., and aluminium cells are used for higher speeds (Cole *et al*., 2008). The use of specialised materials ensures that the cell is able to withstand high gravitational forces, does not leak or change shape. Sedimentation velocity is observed through a column height of 12 mm, with about 400 µ solution in each sector of the cells. Sedimentation equilibrium can employ six sector cells with three 100 µl samples and three reference buffers at 100 µl volume each. The equilibrium distribution is usually observed across a column height of 2 mm or less (Figure 2.2a; Cole *et al*., 2008). The use of shorter cells enables the quicker attainment of equilibrium.

(2.5.2) The optical systems

The optical systems track changes in the distribution of protein molecules during the spinning of the rotor. There are three commercially available optical detectors for the XL-I

(a) Cells and Rotor



(b) Assembly of a cell

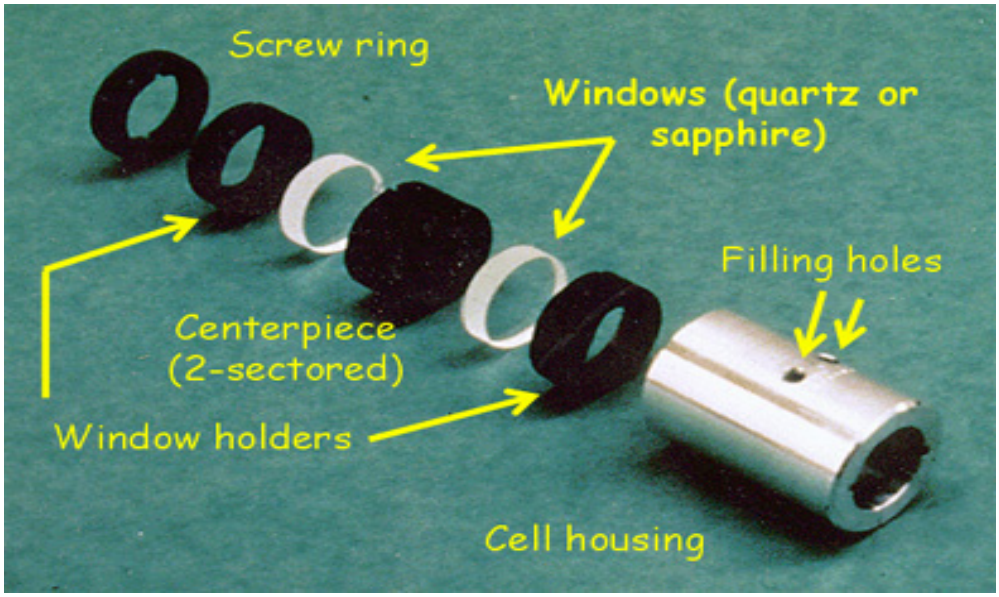


Figure 2.2 Cells and rotor. (a) For sedimentation velocity experiment, 2-sectored centrepieces are inserted into rotor; while for sedimentation equilibrium experiments, 6-sectored centrepieces are usually inserted. (b) Assembly of a cell from different pieces.

to measure the concentration distributions: an absorbance spectrophotometer and Rayleigh interferometer from Beckman Coulter (Palo Alto USA)and the fluorescence detector from Aviv Biomedical (Palo Alto USA). All subsequent analysis of sedimentation data relies on the quality of data available from these detectors. The capabilities and characteristics of the three optical systems are compared in Table 2.1.

The absorbance optical system of the AUC makes use of a high intensity xenon flash lamp and a scanning monochromator. The monochromator measures at wavelengths from 190 to 800 nm (see, e.g Giebeler, 1992). According to Figure 2.3, the light from the Xenon flash lamp passes through the sample and reference sectors and then is detected by a photomultiplier tube. The photomultiplier detects the difference between these signals and translates this into an absorbance value. When the xenon flash lamp moves radically along the length of the sector, different positions of the solute can be scanned. The distribution of protein molecules is constantly changing during the experiment, so the light intensities through the sample sector and reference sector vary as a function of time. Timing information is provided with the aid of a reference magnet at the base of the rotor. Of three optical systems, the absorbance system requires the longest time to complete a scan. For sedimentation velocity experiments, the long scan times may limit the amount of data that can acquired over the course of an experiment.

Interference optics (Figure 2.4) is an optical system of choice that is available to characterise very concentrated or very dilute the samples, due to its sensitivity. It is based on the refraction of light as it moves from one medium to another. The signal from the sample sector is referenced against that of the buffer sector. If the sample and reference solution are the same, refractive index differences will not appear. When the two solutions are different in protein concentration, the achromatic fringe will shift, reflecting the difference in the refractive indices of the solution. During sedimentation experiments, the camera sends the data in digital form to a computer and this will result in a screen image of fringe pattern (Figure 2.5). Because the signal does not depend on chromophores, colourless compounds, such as polysaccharides and lipids, may be analysed by AUC. Since any pathlength difference between the sample and reference beams contributes to the fringe displacement, even tiny optical imperfections (dust, oil or scratches on the lenses) are visible in the signal; cells for interference optics have to be cleaned thoroughly. Also, in order to achieve the

	Absorbance	Interference	Fluorescence
Sensitivity[a]	0.1 OD	0.1 mg/ml	100 pM
Range[b]	2-3 logs	3-4 logs	6-8 logs
Precision[c]	Good	Excellent	Good
Radial resolution[d]	20-50	10	20-50
scan time[e]	60-300	1-10	60-90
When to use[f]	Selectivity Sensitivity Non-dialysable components	Solvent absorbs light Solute does not absorb light Accuracy needed Short solution columns	Selectivity Sensitivity Small sample quantities Non-dialyzable components

[a] The sensitivity is the minimum amount of signal needed to obtain good results. Sensitivity of the fluorescence system is for fluorescein (molar extinction coefficient ~65,000 at 488 nm, quantum yield ~0.9)

[b] The range refers to the concentration range

[c] The precision of the optical system is estimated by comparing the signal to noise ratio.

[d] Approximate spacing (in μm) between data points can be considered, independent estimate of the concentration

[e] The minimum time (in seconds) required to complete one radial scan. The time listed for the fluorescence system is the time needed to scan all of the samples.

[f] Selectivity refers to the ability of absorbance and fluorescence systems to discriminate between components based on their spectral properties. Since the Rayleigh interference optics relies on differences in the refractive index of the sample and reference solutions, it provides no selectivity. The interference optics require that samples are at dialysis equilibrium with the reference solution: hence, they should not be used for samples containing non-dialyzable component. The greater radial resolution of the interference optics allows them to be used with the eight channel "short column" centrepieces.

Table 2.1 Capabilities of optical systems (adapted from Cole *et al*., 2008).

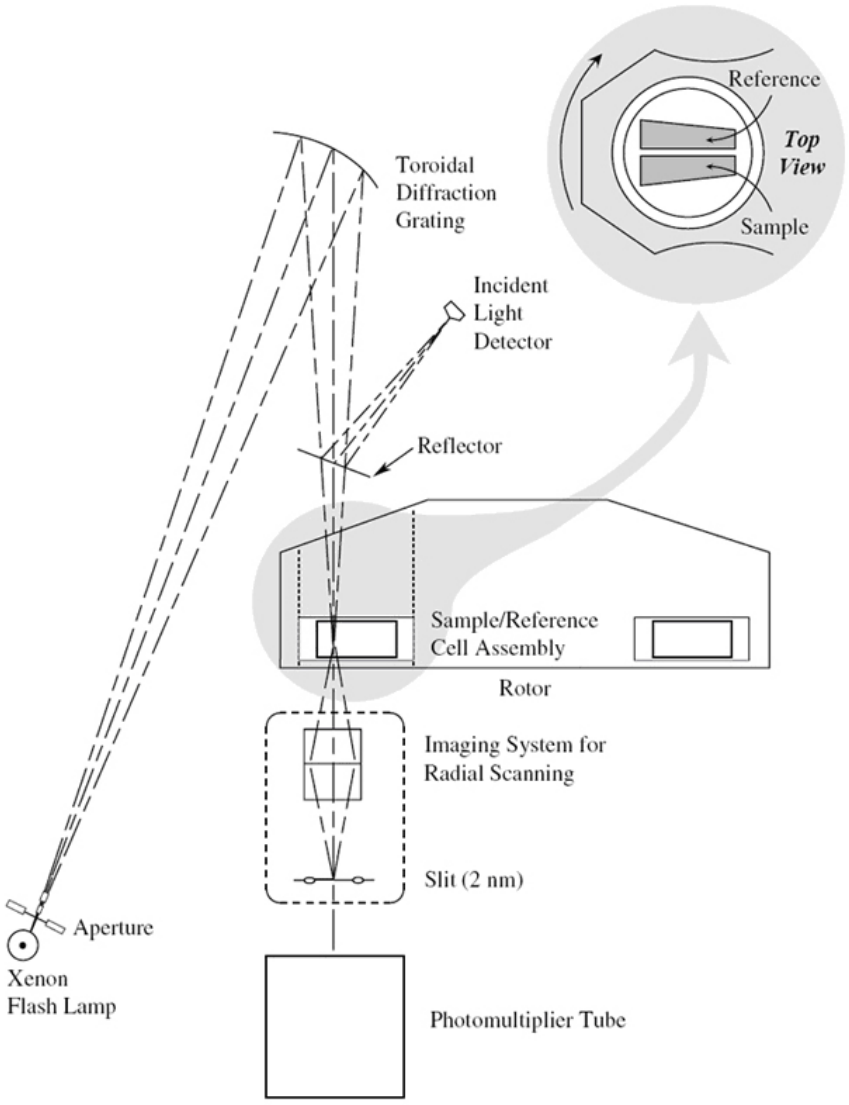


Figure 2.3 A schematic diagram of an analytical ultracentrifuge absorbance optical system (Ralston, 1993).

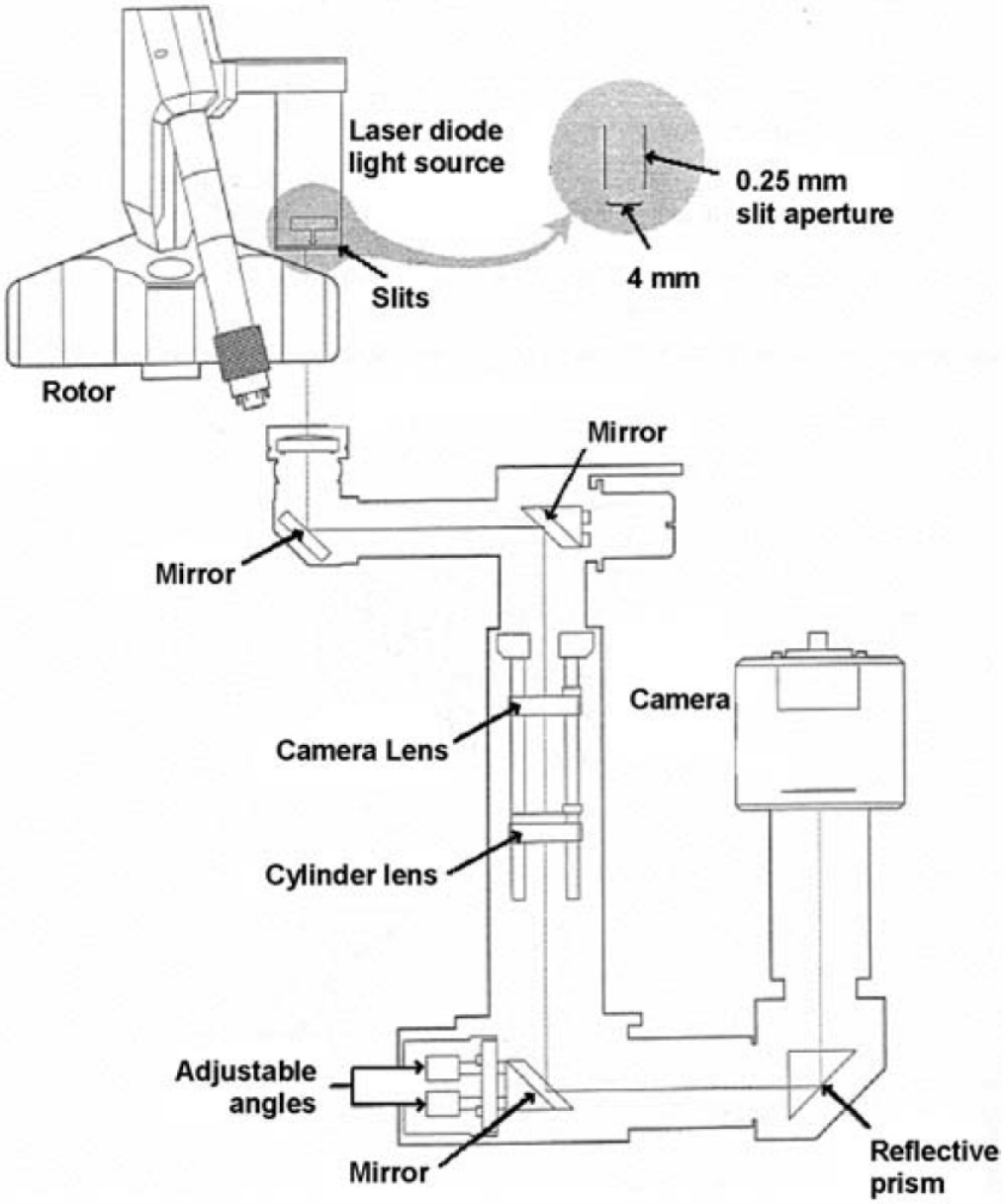


Figure 2.4 A schematic diagram of an analytical ultracentrifuge interference optical system (Beckman XL-I/XL-A Training Guide, 1998).

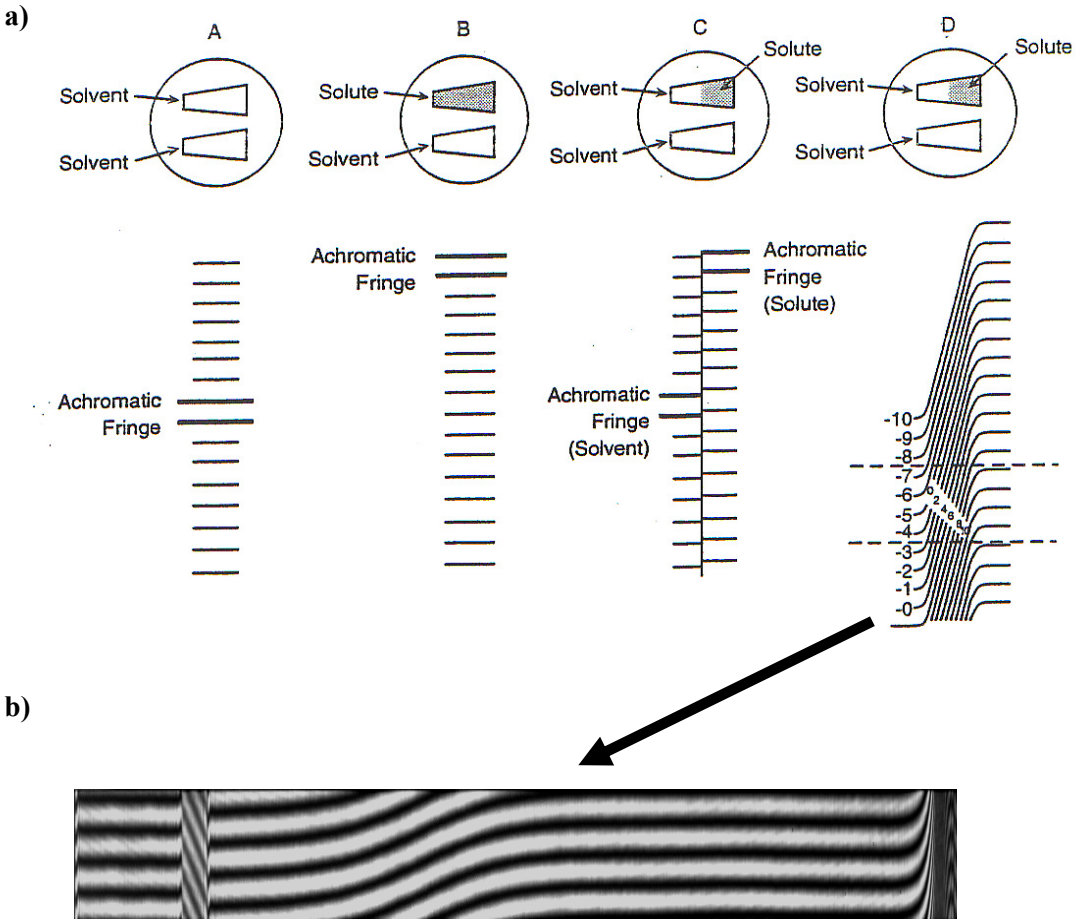


Figure 2.5 Schematic and bitmap of fringe shifts

a) Fringe shifts observed in interference optics of velocity cells. A has no refraction of light. The fringe forms after passing through the same solvent. B shows the achromatic fringe shift due to the difference in refractive indices of the solvent and solute. C shows without diffusion, achromatic fringe shifts; however, XL-I is not able to detect the change. D shows clear fringe pattern as significant amounts of diffusion occur.

b) Formation of achromatic fringe (Beckman XL-I/XL-A Training Guide, 1998).

full accuracy of the interference optics, careful alignment and focusing are necessary (Richards *et al*., 1971). Once properly aligned and focused, they remain stable.

The fluorescence optical system is the most recent addition to the XL-I. A laser light source is used to achieve sufficient radial resolution. Currently the AU-FDS laser provides excitation at 488 nm. It is likely that more excitation wavelengths will become available in the future. The system uses extrinsically labelled compounds. Suitable labels include fluorescein, BODIPY, NBD, green fluorescent protein (GFP), and the many derivatives of these labels used for fluorescence microscopy. Due to the sensitivity and selectivity of fluorescent system, it is possible to characterise the sedimentation behaviour of GFP-labelled protein in cell lysates without further purification.

(2.6) Sedimentation velocity

Sedimentation velocity observes the movement of solute away from the meniscus (narrower end of Figure 2.2a) toward the bottom of the cell. The sample is initially a uniform solution (the concentration is constant over the length of the cell). During centrifugation, the particles move along gravitational force lines, the meniscus is cleared of solutes, and a moving boundary forms between the solvent depleted of solutes and the solvent of finite concentration. The different mobilities of macromolecular solutes in a gravitational field are exploited to separate the species and have different sedimentation coefficient; however, two components of a system may have similar sedimentation coefficient so that they can not be easily separated.

The simplest analysis of a sedimentation velocity experiment was done by plotting the natural logarithm of the radial position of the boundary mid-point or the second moment boundary position versus time (Hansen *et al.*, 1994; Schachman, 1959). The slope of this plot is proportional to $\omega^2 s$, in which s is the average sedimentation coefficient. The methods assume the cells contain a single, noninteracting component.

Later softwares were developed to analyse the entire boundary. One of them is the time derivative approach (DCDT). It implements the dc/dt method originally developed by Walter Stafford, and converts the boundaries into an integral distribution of s, $g(s^*)$. The

$g(s^*)$ distributions are obtained by subtracting pairs of scans, collected over a short time, to generate a set of *dc/dt* data. The differences are then normalised and averaged to enhance the signal-noise ratio and subtract time-invariant baseline distortions (Stafford, 1992a, 1994a). The results obtained by this method are plotted as $g(s^*)$ against s^*. Individual peaks in the $g(s^*)$ distribution can be fitted to Gaussian functions to derive the concentration (from the peak area), the sedimentation coefficient (from the centre position) and the diffusion coefficient (from the width of the peak) for homogeneous species or non-interacting mixtures (Philo, 2006).

In this thesis, another approach is used to analyse the sedimentation boundaries, namely SEDFIT software (Schuck, 1998, 2000). It can load all the scans and test these for evidence of polydispersity, which is different from the DCDT method. SEDFIT fits the boundaries using the Lamm equation (Lamm, 1929), which is derived by combining the law of diffusion and centrifugal migration. The $c(s)$ analysis (Figure 2.6(a,b)) by SEDFIT take into account the diffusion of all species by assuming the same frictional ratio for all species. This assumption is based on the fact that frictional ratio f/f_O is not very sensitive to shape changes, and normally is similar for analogous molecules. The $c(s)$ analysis results in $c(s)$ distribution(s) (peaks) that reflect the populations of different sedimenting species. The integration of $c(s)$ peaks gives the concentration of the species. If a peak corresponds to a single sedimenting species, the $c(s)$ distribution can be transformed to a molar mass distribution $c(M)$. A $c(M)$ distribution presents the approximate molecular weight of the species. Whether the model adequately describes the experimental data is evaluated by goodness-of-fit analyses. The goodness-of-fit in SEDFIT is presented using the residual plot, the grey bitmap image and the r.m.s.d. value. A good fit is shown by a monochromatic grey bitmap image and a low r.m.s.d value (< 0.009).

The traditional approach to analyzing an interacting system is to measure the weight average sedimentation coefficient, $s_{20,w,}$ as a function of loading concentration (Correia, 2000; Correia *et al*., 2005; Schuck, 2003). However, the analysis of reversible interactions by sedimentation velocity is a complex problem (Dam and Schuck, 2005; Dam *et al*., 2005; Schuck, 2003; Stafford, 2000; Stafford & Sherwood, 2004). For interacting molecules, the peaks cannot be assigned to discrete species, because the species compositions are continuously varying due to an equilibrium between them (Dam & Schuck, 2005; Brown & Schuck, 2006). Such a mass-action equilibrium can be seen by the peak position shifting to

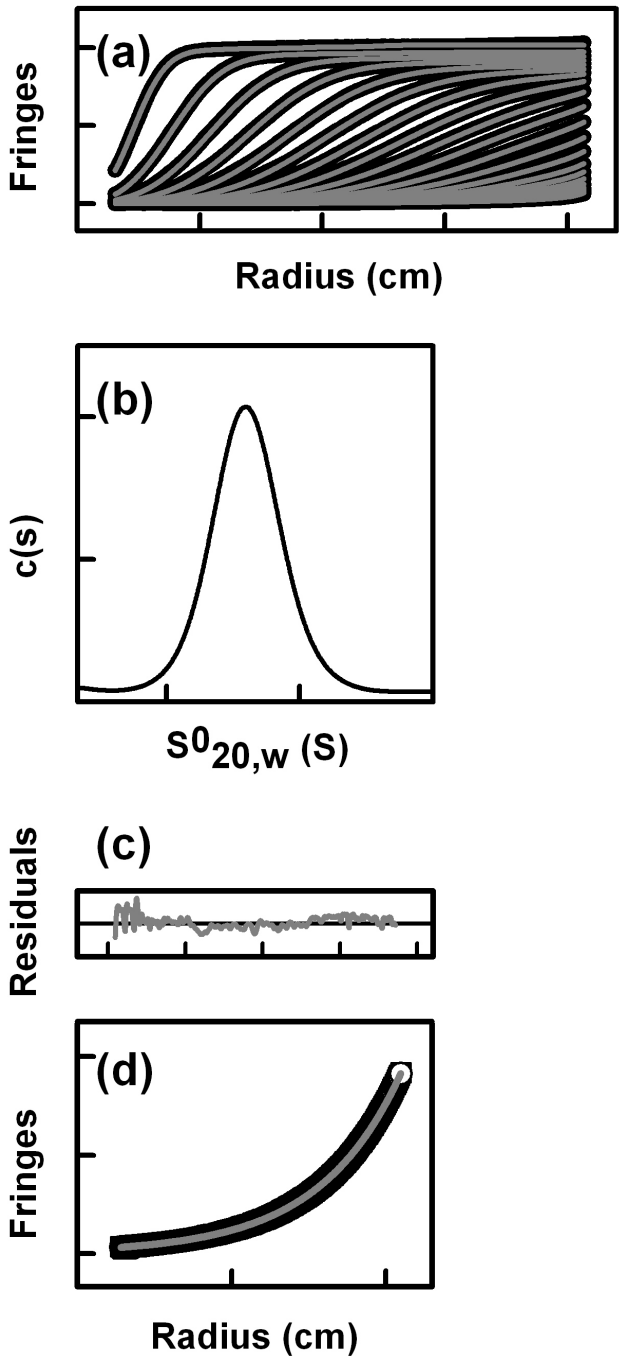


Figure 2.6 Analyses of sedimentation velocity and sedimentation equilibrium by SEDFIT and SEPHAT. (a) The boundary movement of a sedimentation velocity experiment (b) A c(s) distribution derived from the boundary analysis. (c-d) The analysis of sedimentation equilibrium: (c) is the goodness-of-fit shown in residual; (d) is the raw data (circle) and fit of the data.

higher s^* with increasing concentration. There, the peak represents a reaction boundary and cannot be treated as an individual species. Reaction boundaries are peaks that correspond to the co-sedimentation of both free and complexes (Scott & Schuck, 2005). However, if the peaks exhibit moderate or higher stability, the peak positions can give valuable insights into the reaction scheme and stoichiometry.

(2.7) Sedimentation equilibrium

Sedimentation equilibrium (SE) reveals the mass of the macromolecule rather than its shape, because the equilibrium distribution depends on the buoyant molecular weight and not on the shape of the component in solute. The technique is performed at lower speeds and the solute molecules do not completely sediment to the base of the cell, due to the diffusion force (Figure 2.7). Increased sedimentation is balanced at equilibrium by the reverse flow from diffusion. Thus the rate of sedimentation equals the rate of diffusion, and the solute is apparently immobile at equilibrium.

At equilibrium, the concentration distribution is exponential with the square of the radial distribution. In polydisperse solution, higher molecular masses dominated towards the bottom of the cell, while the lower molecular masses dominate at the top. Providing that the values of buffer density ρ and the macromolecular partial specific volume $\bar{v}$ are known, the molecular weight *Mr* can be calculated from the concentration gradient. Standard curve-fit methods using single or multiple species exponential function can provide the molecular weight *Mr* for the solute.

For a single sample, the sedimentation equilibrium needs to be performed at multiple speeds and different concentrations of the solute (Figure 2.8). These are used to find out the optical speed and concentration for calculation of the molecular weight (Perkins *et al*., 2005). The difference between the experimental data points and the corresponding points on the fitted curve which are calculated from the model equation is called the residual. The residual is a function of radius. In an accurate presentation, the residuals should be randomly distributed around a zero basline (McRorie & Voelker, 1993; Beckman Model XL-A/XL-I Analytical Ultracentrifuge Optima™ Series Training Guide, 1998).

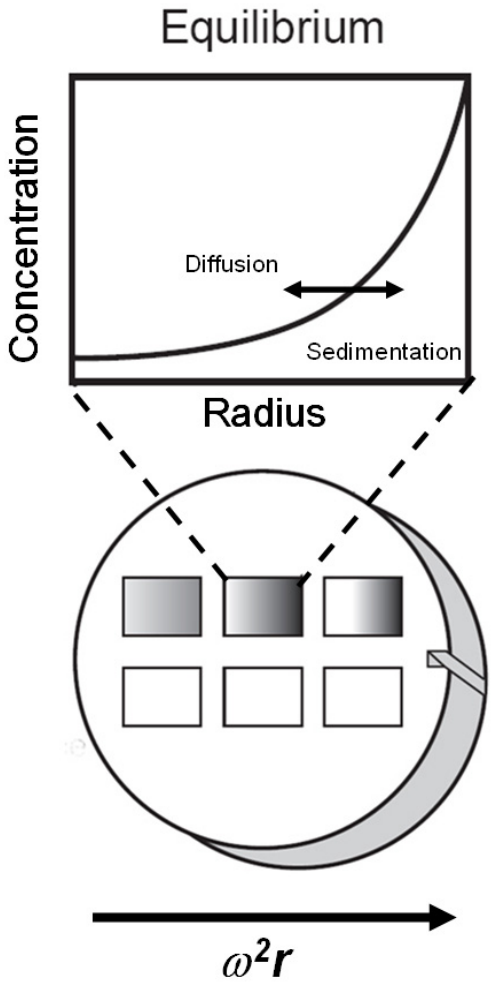


Figure 2.7 Movement of the boundary in a six-channel centrifuge cell during sedimentation equilibrium (Cole & Hansen, 1999). SE data are collected when the protein has reached equilibrium between its diffusion and sedimentation in the cell.

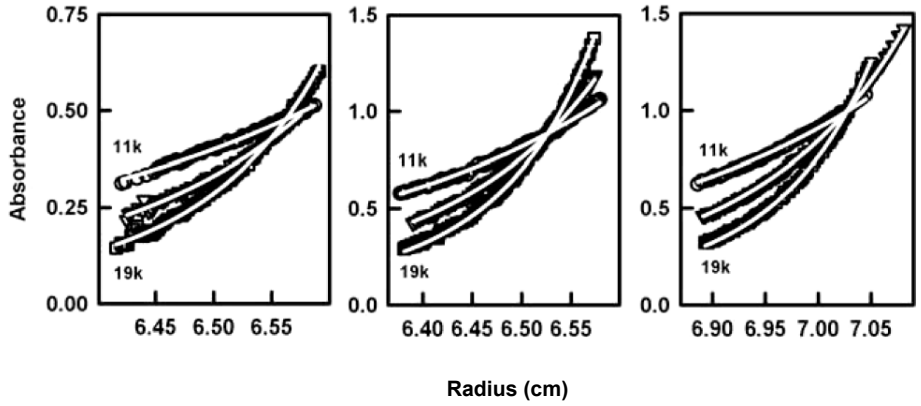


Figure 2.8 Sedimentation equilibrium analysis of the single-chain Fv fragment MFE-23. The curves were obtained at three concentrations of 1.1, 2.2 and 3.0 mg/ml with three speeds of 11, 000, 15, 000 and 19, 000 r.p.m.. The continuous white lines are the result of the simultaneous fit using non-linear least-squares regression analysis (Perkins *et al*., 2005).

For an ideal non-interacting simple component system, the following equation applies:

(Eq. 2.10) $c_r = c_{r0} \exp[Mr\,(1-\bar{\upsilon}\rho)\,\omega^2\,(r^2 - r_0^2)/2RT]$

Where c_r is the sample concentration at radial position r, c_{r0} is the concentration at meniscus, Mr is the molecular mass of solute, ρ is the density of the solvent, ω is the angular velocity, r is the radial distance from the meniscus, R is the gas constant (equal to 8.314×10^7 erg $K^{-1}mol^{-1}$) and T is the absolute temperature (Cole & Hansen, 1999; Beckman XL-I/XL-A Training Guide, 1998). Equation 2.10 is not affected by the molecular shape. The shape of molecules only affects the rate at which equilibrium is reached. An example of single species fit by Equation 2.10 is in Figure 2.6(c-d).

Sedimentation equilibrium can also be used to characterise oligomerisation of a molecules. For self-association systems, the total absorbance at a given radius is the sum of absorbance of all species at that radius, and the total concentration distribution can described as sum of two species:

(Eq. 2. 11)

$$c_{total}(r) = c_m(a) \exp[Mr\,(1 - \bar{\upsilon}\rho)\,w^2\,(r^2 - a^2)/2RT] + c_n(a) \exp[nMr\,(1 - \bar{\upsilon}\rho)\,w^2\,(r^2 - a^2)/2RT]$$

where $c_{total}(r)$ is the total concentration of the mixture at the radial position, $c_m(a)$ is the concentration of monomer at the meniscus, $c_n(a)$ is the concentration of oligomer at the meniscus, and n is the stoichiometry for the oligomer. Each exponential in the summation describes the equilibrium distribution of a species, the first being the monomer and the second the n-mer. For the monomer-oligomer equilibrium reaction, the association constant, K_a, is defined as: $K_a = c_n / (c_m)^n$ (McRorie & Voelker, 1993).

Although the detailed mathematical theories that underlie sedimentation velocity and sedimentation equilibrium are complex, many user-friendly data analysis programs enable users, with little experience of computer analytical technique, to analyse their AUC data with confidence (Table 2.2). In my thesis, I used SEDFIT to analyse the sedimentation velocity experiments and SEDPHAT to analyse the sedimentation equilibrium.

Method	Application	Platform
Sedimentation Velocity		
Van-Holde-Weischet	UltraScan	UNIX, PC
Time derivative	DCDT, DCDT+, UltraScan, Beckman	PC, Macintosh, UNIX
Direct fitting (approximate solution)	SVEDBERG LAMM	PC PC
Direct fitting (numerical solution)	UltraScan SEDFIT	UNIX, PC PC
Second moment	UltraScan, Beckman	UNIX, PC
Hydrodynamic modeling	HYDRO ATOB	FORTRAN Source Code PC
Sedimentation Equilibrium		
Nonlinear least squares	NONLIN, Beckman UltraScan SEDEQ	PC, Macintosh, VMS UNIX, PC PC
Hetero-association	TWOCOMP	PC
Ω function	SEDProg	PC

Table 2.2. Some of the software for analysis of sedimentation data (Cole & Hansen, 1999).

(2.8) Non-ideality in SV and SE experiments

A solution is considered to be ideal when its enthalpy is zero; the closer to zero the enthalpy of solution is, the more "ideal" the solution behaviour becomes (Pusey & Tough, 1985). Deviations from ideal behaviour are corresponding to non-ideality. In a non-ideal solution, there is a tendency of a suspension of particles to remain in solution, to aggregate and to overcome phase separation. In SV experiments, non-ideality causes the sedimentation coefficient and diffusion coefficient values to vary with concentration. In SE experiments, non-ideality causes variation in the point average molecular weight depending on concentration or radial position. In this case, the measurement of molecular weight by SE over a range of concentrations is necessary followed by an extrapolation to zero concentration (Harding & Johnson, 1985). Biological macromolecules in concentrated solutions reflect the effect of macromolecular crowding on biological function (see, e.g., Harding & Johnson, 1985).

In SV experiments, sedimentation coefficients from concentration series, s, are corrected to be that measured at a temperature of 20°C and to water (solvent environment), and labelled as $s^0_{20,w}$ (Laue *et al*., 1992). The variation in the sedimentation coefficient with concentration may be caused by an increment in the frictional coefficient with concentration. The concentration dependence of the diffusion coefficient arises from both the thermodynamic nonideality and the hydrodynamic nonideality (Solovyova *et al*., 2001). The driving force for diffusion is the gradient in chemical potential (μ), which is related to osmotic pressure (Harding & Johnson, 1985). The shapes of the sedimentation profiles generally show steeper boundaries for repulsive non-ideal sedimentation. In SV experiments,

(Eq. 2.12) $$c(r,t) = f\,(f/fo, Mr, k_S, K_D, r, t\,);$$

where k_S is a mass-action parameter that determines hydrodynamic non-ideality. At high protein concentrations, repulsive interactions between solute molecules cause a reduction in the sedimentation coefficient, which can be expressed by the non-ideality coefficient for sedimentation k_S (an empirical constant in ml/g) (Schachman, 1959). A single peak distribution seen in c(s) plots does not always mean a single species. Normally low protein concentrations under 1 mg/ml are used to avoid non-ideality. The salt concentration in buffer should be at least 20-50 mM to shield electrostatic interactions between macromolecules that contribute to thermodynamic nonideality (Cole *et al*., 2008).

Thermodynamic non-ideality affects the spatial distribution of macromolecular solutes at sedimentation equilibrium. The point average molecular weight depends on the concentration or radial position (Wills & Winzor, 1992). Measurement of the relative molecular mass of a macromolecule over a concentration range is necessary followed by an extrapolation to zero concentration (Harding & Johnson, 1985). The effects of non-ideality can be summed up in terms of a "virial expansion" (see, e.g., Scott *et al*., 2005). In SE experiments,

(Eq. 2.13) $c(r) = f(Mr, B, K_D, r)$;

the second virial coefficient B (the first coefficient is $1/M$) explains thermodynamic non-ideality (see, e.g., Harding & Johnson, 1985). Sedimentation equilibrium is able to analyse non-ideality in polydisperse systems. However, the application of theoretical developments to experimental situations is limited by: 1) an inability to accommodate the consequence of reactant heterogeneity; 2) a requirement for valency determination to each interacting species (Winzor *et al*., 2004a, 2004b).

Many macromolecular systems give symmetrical Schlieren peaks in a SV experiment, or constant point-average molecular weights throughout the solution column in a SE experiment. However, such criteria alone are insufficient to prove that a system is ideal and homogeneous (see, e.g., Creeth & Knight, 1964; Teller, 1965).

In the experiments of this thesis, non-ideality does not have a significant effect. Because the experiments have been done in a low concentration range, any concentration effect of protein behaviour becomes standardised when it is converted into $s^0_{20,w}$. Salts have important effects on the stability and solubility of macromolecules in solution. The concentration of salt and valency of ions are only taken into account, when the solution has very low (millimolar) salt concentrations (Serdyuk *et al*., 2007). My experiments have been done in both 50 mM NaCl and 137 mM NaCl buffers. The use of 50 mM NaCl increase the electrostatic interaction between molecules compared to that of 137 mM NaCl. It is possible that the results in 50 mM NaCl buffer induced artificial inter-particle interaction. My AUC experiments were validated by X-ray experiments to confirm that even in low salt, my results show no notable effects arising from non-ideality (Chapter 4, Chapter 5 and Chapter 6).

Chapter Three

Solution Scattering

(3.1) Why use solution scattering?

Solution scattering is a low resolution diffraction technique that is used to study the overall structure (length, width and conformation) of biological macromolecules in random orientations (Perkins *et al*., 2008). This method is used for proteins with a molecular mass of 10 kDa or upward, with no upper limit. The major strengths of solution scattering are that it provides multi-parameter structural information on proteins that are analysed in near-physiological conditions. X-ray crystallography is able to obtain protein structure at atomic resolutions of up to 0.1 nm, but only fixed orientation view. In contrast, solution scattering allows structures to be viewed at resolutions of about 2-4 nm, but in various orientations. Analytical ultracentrifugation provides structural information through sedimentation or diffusion coefficients. This is a single-parameter method, while the related method of solution scattering is a multi-parameter method. Solution scattering can analyse multi-domain protein structures that are not crystallisable for various reasons, such as heavy glycosylation or the presence of flexible inter-domain linkers. The concentration of proteins required are lower than many other techniques, which minimises the need for non-physiological high concentrations of sample. Furthermore, solution scattering can be combined with X-ray crystallography to validate protein crystal structures in solution, and to identify any significant oligomeric or conformational differences between the solution and crystal structure (Perkins *et al*., 2009). Overall, solution scattering overcomes some of the limits of other methods of structure determination. It can describe a large macromolecule structure in near-physiological condition in less time and model the scattering data quantitatively by using known atomic structures. Therefore, solution scattering is a useful complementary tool in the research area of protein structures.

(3.2) Types of solution scattering

There are two common variants of solution scattering: X-ray scattering and neutron scattering (diffraction by electrons and nuclei respectively). They share the same physical principle. A sample solution is irradiated by a highly collimated (divergent or convergent rays converted into parallel rays with minimum divergence or convergence), monochromatic (single wavelength) beam of X-rays or neutrons. The diffraction pattern is then recorded on an area detector at low scattering angles. X-ray scattering uses a high degree of collimation of the main beam and a high positive solute-solvent contrast (Table 3.1). Neutron scattering differs by the ability to exchange hydrogen by deuterium in the

solvent water so that the appearance of the macromolecule can be varied by the contrast variation method; and different parts of the sample can be seen in different contrasts. This is valuable for protein-DNA complexes. By X-rays, the effect of incoherent scattering (where the scattering is out of phase) is negligible; however, it becomes a more noticeable issue by neutron scattering. A background level of uniformly scattered neutrons, especially at large scattering angles, is created by incoherent scattering. Incoherent scattering from hydrogen in the protein requires a flat-background correction to solve the problem in neutron scattering analysis (Perkins *et al*., 2008).

(3.3) Solution scattering experiments

In a scattering experiment, a solution of the sample is irradiated by a highly collimated beam of X-rays (wavelength $\lambda \approx 0.1$ nm) or neutrons ($\lambda \approx 0.5$ nm). As a result of coherent scattering and good contrast between the protein and buffer, a circularly-symmetric diffraction pattern is recorded on a two-dimensional area detector that is placed behind the sample. The scattered intensities *I(Q)* are measured as a function of scattering vector Q, where $Q = 4\pi.sin\theta/\lambda$ (2θ is the angle between the incident and scattered radiation, λ = wavelength) (Perkins, 2002; Figure 3.1). The resolution of a scattering experiment is usually defined by $2\pi/Q_{max}$, where Q_{max} is the largest Q value measured in the experiment. (Svergun & Koch, 2003). A typical solution scattering pattern is obtained in a Q range between about 0.05 nm^{-1} to over 2 nm^{-1}. The lowest Q values are not measured due to the beam stop to protect the detector camera from the high intensity of the direct main beam. Data acquisition times range from seconds to hours. After data have been collected, the *I(Q)* data are analysed by Guinier plots at low Q to extract the R_G and *Rxs* parameters and molecular weights. Data are also analysed by indirect transformation to obtain the distance distribution function $P(r)$. After satisfactory data analysis, constrained modelling calculations are performed to verify and extend the data interpretation in terms of structures.

(3.3.1) The Debye equation

Diffraction is described by Bragg's Law according to the equation

(Eq. 3.1) $$\lambda = 2d\ sin\theta,$$

The equation explains the wavelength (λ) in terms of the diffraction spacing (d) and

Table 3.1 Comparison between X-ray and neutron scattering methods.

X-ray	Neutron
Studied in high positive solute-solvent contrasts; a single contrast measurement	Contrast variation allow different regions within proteins or complexes to be seen
Instrumental errors caused by wavelength polychromicity and beam divergence are minimal, no correction required	Affected by beam divergence, wavelength spread and a high uniform incoherent scattering background; needs correction
Radiation damage effects need control ; sample cannot be recover after irradiation	No radiation damage effects; sample can be reused
A monolayer of water molecules is hydrogen bonded to the protein surface. The macromolecule appears larger	No hydration shell is visible, structures corresponds to water-free protein coordinates
Leads to relative *Mr* values	Leads to absolute *Mr* values
Buffer background are affected by high salt	Buffer background are not affected by high salt
Sample aggregate after radiation damage	Sample may aggregate in a non-specific manner in heavy water solvents

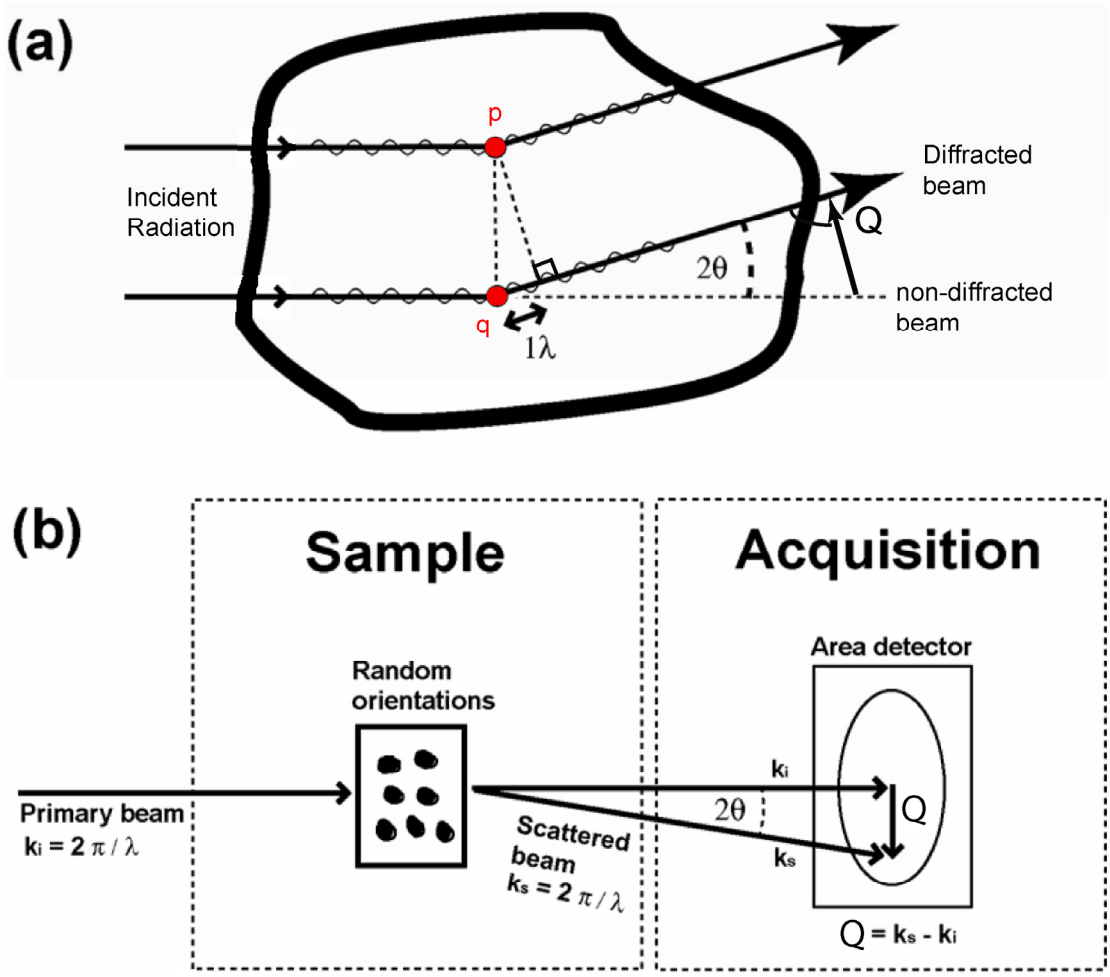


Figure 3.1 Schematic representation of a scattering experiment.

(a) Scattering from two points p and q in a protein molecule are shown (red circles). The scattered angle 2θ and scattering vector *Q* are also shown. (b) The sample in random orientations is irradiation by primary beam and then recorded by an area detector. The preliminary image is a scattering pattern about the position of the direct main beam (masked by a beam stop). k_i is the incident beam, while k_S is scattered beam (Perkins *et al.*, 2008).

scattering angle (2θ). In solution scattering, d corresponds to dimensions of 1-100 nm. Scattering angles are given in terms of the scattering vector Q (Section 3.3). For any pair of scattering electrons or nuclei, if 2θ is equal to zero, the scattered waves are in phase and the scattering intensity is at its maximum. In such a situation, the intensity of scattering, $I(0)$, is a direct measure of molecular mass, when $Q = 0$ means at angle zero. When Q and 2θ are not equal to zero, the intensity of scattering $I(Q)$ is described by the Debye equation (Debye, 1915; Guiner & Fournet, 1955):

(Eq. 3.2) $$I(Q) = \sum_{p}\sum_{q} f_p f_q \frac{\sin(rQ)}{rQ}$$

where the terms f_p and f_q (or b_p and b_q for neutrons) are the scattering lengths of the electrons (or nuclei; Section 3.3.2) at points p and q in the molecule, and r corresponds to the distance between p and q. The scattering length, f, is proportional to the atomic number for X-rays, while b for neutrons is roughly similar for different nuclei with the exception of ^{1}H (Table 3.2a). In solution scattering, macromolecular structures are described in terms of scattering power of atoms and the geometrical relationship between the individual atoms within the sample. The Debye equation takes account of both factors (Perkins, 1988). A limitation of the Debye equation is that it describes particles *in vacuo* and assumes that the scattering patterns are circularly symmetrical. Therefore, the Debye equation is justified for two-phase model of aqueous proteins randomised in a solution.

(3.3.2) Contrast variation

Only when the scattering density of the protein is different from the scattering density of the dissolved buffer, can the protein be observed by solution scattering. The scattering density of a solute or a buffer is the sum of atomic scattering lengths divided by molecular volume V. X-ray scattering is based on electrons. The scattering length f of an electron is 2.81 fm (or 2.81×10^{-15} m, the classical electron radius). The specific scattering lengths for different materials are listed in Table 3.2a. The hydrogen isotopes, ^{1}H and ^{2}H, both have the scattering length of an electron of 2.81 fm.

For neutron scattering, b represents the atomic scattering length. It is different from X-ray scattering, because neutrons interact directly with the nucleus of the atom. The nuclei

(a)

		Atomic number	$f(2\theta = 0^o)$ (fm)	b (fm)
Hydrogen	^{1}H	1	2.81	-3.742
	^{2}H	1	2.81	6.771
Carbon	^{12}C	6	16.9	6.651
Nitrogen	^{14}N	7	19.7	9.40
Oxygen	^{16}O	8	22.5	5.804
Phosphorus	^{31}P	15	42.3	5.1

(b)

Macromolecule	X-rays ($e.nm^{-3}$)	Neutrons (% $^{2}H_2O$)
H_2O	334	0
$^{2}H_2O$	334	100
50 % (w/w) sucrose in H_2O	402	13
Lipids	310-340	10-14
Detergents	300-430	6-23
Proteins	410-450	40-45
Carbohydrates	490	47
DNA	590	65
RNA	600	72

Table 3.2 Scattering lengths and densities for different materials (Perkins, 1988).
(a) Scattering lengths of biologically important nuclei. (b) Scattering densities of solvents and biological macromolecules. "$e.nm^{-3}$" is the unit for electron density. "% $^{2}H_2O$" is neutron contrast difference between varies materials and 100% heavy water.

of all atoms are compact and of comparable size, so neutrons are capable of interacting strongly with all atoms with a similar scattering length. ^{1}H has a single proton in its nucleus, so the scattered intensity is different from that of ^{2}H where there is a proton and a neutron. The atoms ^{2}H, ^{12}C, ^{14}N, ^{16}O, ^{31}P and ^{32}S all have comparable positive b values. (Table 3.2a). After dividing the scattering lengths by the molecular volume, it is found that the protein scattering density in $^{2}H_2O$ corresponds to 40-45% $^{2}H_2O$, based on a scale ranging from 0% to 100% $^{2}H_2O$ (Table 3.2b). The large difference in the neutron scattering properties of the two hydrogen isotopes can be utilised in contrast variation neutron scattering experiments.

Solution scattering is a low resolution technique (1-4 nm). Usually, atomic scattering lengths in the Debye equation can be replaced by scattering densities $\rho(r)$ of small volume elements. The Debye equation is modified to consider the contrast $\Delta\rho$, which is the difference between the mean scattering density of the macromolecule ρ_v and that of its buffer ρ_s:

(Eq. 3.3) $$\Delta\rho = \rho_v - \rho_s$$

The terms ρ_v and ρ_s are calculated from the total scattering lengths Σf_p or Σb_p divided by the volume V (Eq. 3.2; Perkins, 1988). Larger $\Delta\rho$ values give rise to larger signals, and this is important for scattering from dilute solutions. Subtracting the scattering of the buffer from the sample must be done as precisely as possible to accurately measure differences of over three orders of magnitude as Q increases (Putnam *et al*., 2007).

Scattering density may vary within biological molecules. Proteins, carbohydrates and nucleic acids have progressively fewer H atoms and greater proportions of N and O atoms; in that order thus their electron densities are greater than that of water; whereas lipids possess a great proportion of H and C atoms which are less electron-dense than H_2O molecules. Table 3.2b shows that neutrons have a wider range of solvent contrasts than X-rays. Neutron scattering can discriminate between the four macromolecular classes, lipids, proteins, carbohydrates and nucleic acids.

Overall, contrast variation is important in both X-ray and neutron scattering. For X-rays, the protein scattering density is distinguishable from the buffer scattering as it has a higher scattering density (positive contrast). For neutrons, the major problem with the

contrast is to find suitable solvents with sufficiently different densities that do not alter the structure of the dissolved molecules. Neutron scattering is good at this respect, because large contrast variation is possible by the simple exchange of H_2O by using H_2O and ^2H_2O mixtures, and ^2H_2O is a common solvent.

(3.3.3) Hydration Shell

SAXS experiments analyse hydration effects during data interpretation. Biological molecules have heterogeneous surface chemistry; the interactions between biological surfaces and surrounding water are governed by the dipole moment polarisability and hydrogen-bonding ability of individual water molecules. These interactions modify the structure and dynamics of water near the surface of proteins and a distinct hydration shell is formed (Halle, 2004). Generally the interaction only affects the first hydration shell at about 0-0.3 nm from the protein surface (Raschke, 2006). The average density of the first hydration shell is about 15-22% higher than that of bulk water (Perkins, 2001b). The electron density of the first hydration shell is indistinguishable from those of the amino acids, and it will therefore contribute to the protein X-ray scattering curve (Perkins, 2001b). The mean volume of a water molecule in the hydration shell is 0.0245 nm^3, whilst that of bulk water is 0.0299 nm^3 (Perkins, 1986b, 2001b). The volume occupied by a bound water molecule is ~ 20% smaller than that of free water, causing this to show a higher density (Svergun *et al*., 1998). For SAXS experiments, an assumption of 0.3 g H_2O/g protein leads to a hydration shell volume of approximately 33% of that of the protein.

(3.4) Solution scattering methods

(3.4.1) Sample preparation

Solution scattering experiments require pure and monodisperse samples. The purity of samples can be assessed by reducing and non-reducing SDS-PAGE and AUC. Aggregation is another problem. These lead to increased intesities at low Q which mislead data interpretation, as the scattering intensities at low Q are proportional to the square of the molecular weight. Freezing and thawing cycles promote sample aggregation and are best avoided. A trace amount of very large aggregates (compare to the particle that is being studied) is better than a monomer and oligomer mixture. In the former case, the scattering from aggregates can be removed by removing lowest-q data. In the latter case, the entire

scattering curve is affected and this is not always detectable. AUC (Chapter 2) can be used to detect low levels of aggregation and polydispersity of proteins. Size exclusion gel filtration or centrifugation are used to remove aggregation before data collection.

As mentioned in Section 3.3.2, the greater the difference in scattering densities between the buffer and the protein, the higher the contrast; therefore signal-noise ratio is better. Therefore, buffers for scattering experiments have to be chosen carefully. Salts in buffer increase background and decrease the solute-solvent contrast, but these effects are often negligible up to about 500 mM salt. The use of 1 M NaCl reduces the X-ray transmission to about 10% (Svergun & Koch, 2003). Free radical scavengers sometimes are included in the buffer to minimize radiation damage. Common choices are DTT (2-10 mM), TCEP (1-2 mM), or glycerol (~5%). Organic buffers containing TRIS or HEPES can also act as radical scavengers. Detergents are best avoided unless absolutely necessary, because their scattering could be similar to or exceed the protein scattering. Phosphate buffered saline and HEPES buffer were used for my X-ray scattering experiments. Samples for neutron scattering experiments are prepared similarly as for X-ray scattering experiments, except that PBS buffer in 100% $^{2}H_2O$ was used. Each protein sample is dialysed against the $^{2}H_2O$ PBS buffer. The buffer was changed four to five times to ensure full ^{1}H-^{2}H exchange.

Preparation of an exactly matching buffer is the most crucial step. Long hours (18- 48 hours) dialysis works best. In cases when dialysis cannot be done (limited sample stability or limited time for preparation), the buffer exchange can be done by passing buffer by centrifugation though an appropriate molecular weight cutoff filter. Fresh membrane filter normally have organic chemicals on it, so the membrane needs to be washed with the buffer before the buffer exchange process. Several cycles of buffer loading will be needed to ensure the exact match.

The concentration of sample should be as high as possible to give the best signal, but avoiding aggregation or pronounced inter-particle repulsion. For proteins of *Mr* 200 kDa or less, 1-10 mg/ml is a suitable range. For larger proteins, concentrations should be below 2 mg/ml. Concentration series by dilutions will ensure consistent results. Analysis of $I(0)/c$ (c is the sample concentration in mg/ml) should indicate whether concentration effects are present (such as oligomerisation with increased concentration), because $I(0)/c$ will be

invariant with concentration for monomeric proteins. Protein concentrations are measured from the optical densities at 280 nm in a UV spectrophotometer in quartz cells with known pathlengths. The concentration is then calculated from the Beer-Lambert law, c =A / εl, where c is the protein concentration (mg/ml), A is absorbance at wavelength 280 nm (for protein), ε is the protein extinction coefficient (mg/ml/mm) and l is the pathlength (mm). The extinction coefficients were calculated from its amino acid and carbohydrate composition (Perkins, 1986b). Sample volumes for X-ray work are 90-150 µl for the ESRF flow-cell. When preparing the buffer for X-ray, at least 50 ml of buffer will be needed in order to wash the flow cell before the next sample with the same buffer. Neutron experiments need less buffer, as no washing process is needed, because of the use of sealed quartz cells.

(3.4.2) X-ray data collection on instrument ID02 in ESRF

Biological experiments usually involve dilute monodisperse samples. In order to obtain good signal-noise ratios, high-flux facilities are required for X-ray scattering. X-ray synchrotron radiation sources are more intense than a conventional laboratory anode source. They generate high energy electron beams which are maintained in a storage ring for many hours. There are about 70 X-ray synchrotron sources in the world. The most powerful "third generation" include the 6 GeV European X-ray synchrotron radiation facility (ESRF) in France, the 7.0 GeV Advanced Photon Source in the USA and the 8.0 GeV Super Photon Ring (Spring8) in Japan (Perkins *et al*., 2008). They specialise in either short-wavelength X-rays (high-energy) or vacuum-ultraviolet and long-wavelength X-rays.

The production of synchrotron radiation is based on the laws of electromagnetism. At ESRF, production occurs in several stages. First, electrons are produced by a 100 KeV electron triode gun and are accelerated to 200 MeV in a linear accelerator (LINAC, Figure 3.2(a,b)). Then the electrons are transferred to a 300 m fast-cycling booster synchrotron in which they are accelerated to energies of 6 GeV. Then the electrons are injected into the 844.4 m diameter storage ring to follow a pseudo-circular orbit by bending magnets. The electrons are focused by 320 quadrupoles. There are insertion device sources, undulator and wigglers to force the electron beam onto a sinusoidal trajectory. Because the electrons in the storage ring travel at very high speeds and are constrained in a circular ring trajectory by electromagnets, the electrons radiate high energy, including X-rays. The lost energy is replenished by an oscillatory radiofrequency electric field which maintains the energy at 6.0 GeV.

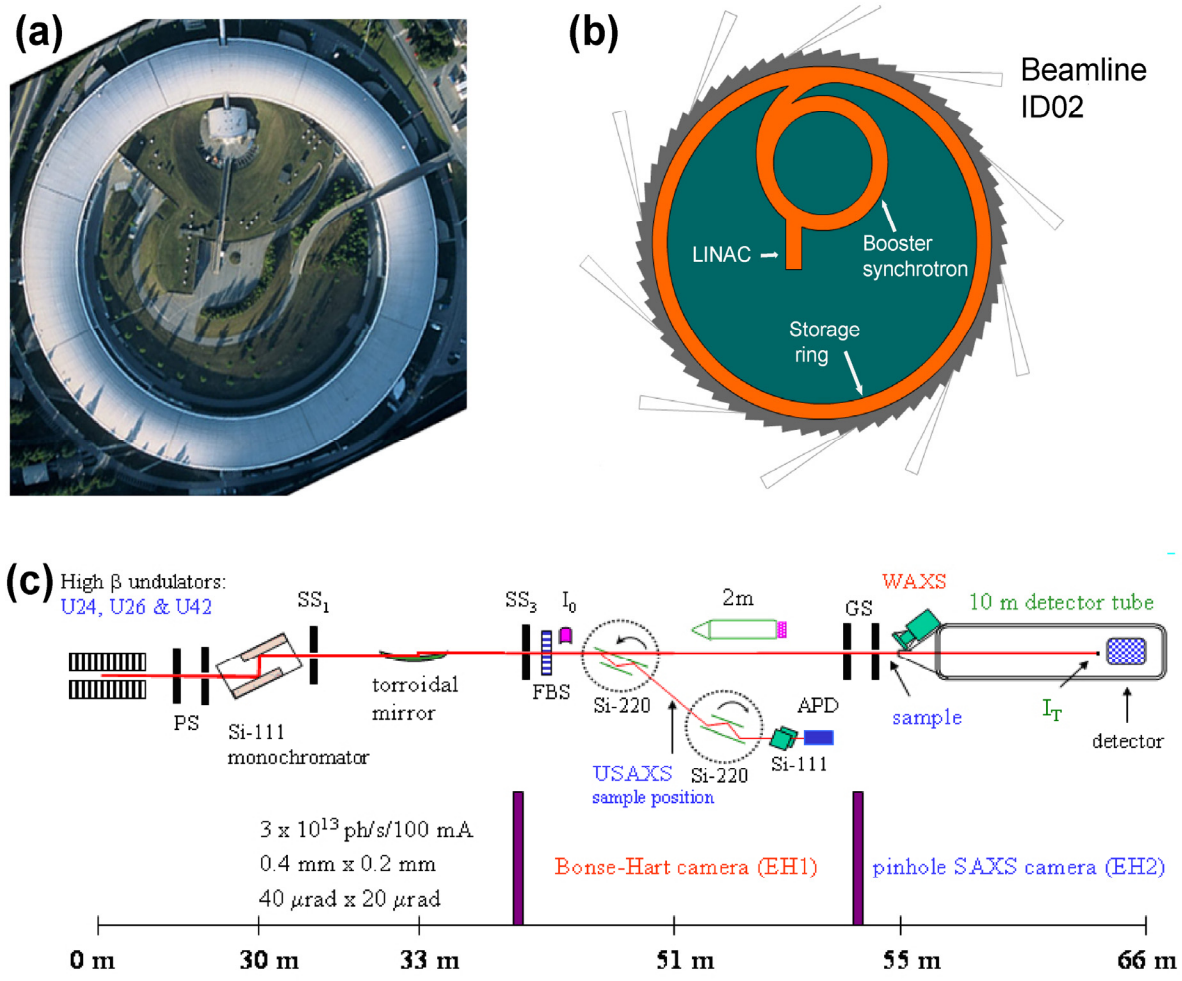


Figure 3.2 Schematic layout of X-ray solution scattering cameras at beamline ID02 at ESRF in France.

(a,b) The source is located at the upper section of the storage ring as shown.

(c) The undulators U24, U26, and U42 provide high photon flux with a low divergence at ID02. Only the central radiation cone of the undulator is used. The beamline operates at a fixed wavelength around 0.1 nm (12.4 keV) by a monochromator-mirror optical system, and the experiments in this thesis utilised the pinhole SAXS camera EH2 (adapted from the ESRF Web site *http://www.esrf.fr*) (PS = primary slits, SS1& SS3 = secondary slits, GS = guard slits).

The ultimate aim of synchrotron radiation is to produce high quality and high brightness beam. Brightness is the measure of how many photons/sec can be imaged on a small sample with a small angular divergence. Higher brightness also means that a better scattering signal will be obtained. It is affect by the flux and amount of emittance (phase space area). For a given wavelength, the total beam power is proportional to the flux. Therefore, high flux and more collimated beam produce higher energy beams onto the experimental samples (Koch *et al*., 2003). Figure 3.2c shows the monochromator and focusing mirror used to make the beam line more efficient for sample radiation.

Electrons circulating around the storage ring emit a white beam of all wavelengths tangential to the ring. Instrument ID02 will receive this, and transform it into a usable X-ray beam for experiments (Narayanan *et al*., 2006; Panine *et al*., 2006). The beamline use a liquid nitrogen cooled Si-111 monochromator and a focusing toroidal mirror (Figure 3.2c) to select a wavelength of 0.1 nm and focus the beam on the sample. There are primary, secondary and guard slits to reduce parasitic background scattering. These parts of the instrument are in a radiation shielded hutch to protect users from lethal X-ray exposure. ID02 is composed of two cameras to monitor three types of scatterings. One is the wide angle scattering detector, immediately after the guard slit, and outside the vacuum of the ten metre long tank within which is a second detector mounted inside the long vacuum tank. Alternatively, there is a Bonse-Hart camera for ultra-low angle scattering (not used in this thesis).

The samples for small angle scattering are placed between the main beam and the detectors (Figure 3.2c). The optimum sample thickness is a trade off between the low concentration of the sample (the less it absorbs light, the less radiation damage is observed), and the more concentrated sample (better scattering intensity). All samples are translated across the beam through a 1 mm wide quartz capillary controlled by a mechanically-operated syringe. This means that fresh sample can be continuously exposed during data collection in order to eliminated radiation damage effects.

Adjustable sample-detector distance (0.5 to 8 m) is used to obtain the desired Q range. The scattered intensity will decrease proportionately to the inverse of the sample-

detector distance squared. It is necessary to optimise a high signal and large Q range data to match the different size of proteins. Two dimensional area detectors in ID02 were used to handle the high scattered intensities. There is an image intensifier device to convert the X-ray photons into visible photons, and then these are recorded by detector (resolution 2048X2048 pixels). The detector is protected from the main beam at zero scattering angle by a beam stop. The detector also connects with computers and workstation for data storage and processing. During the experiments, automated software convert 2D data set to a 1D curve for display; so each sample scattering can be checked quickly to observed and avoid radiation damage.

(3.4.3) Neutron data collection on instruments

There are about 40 neutron sources in the world at moment. Neutron sources include neutron reactor instrumentation and spallation neutron instrumentation. Presently, the most powerful one is the 58 MW High Flux Reactor at the ILL in Grenoble, France. The ISIS facility adjacent to Diamond, Oxfordshire, UK, is one of the successful spallation neutron sources. A neutron camera at a spallation source differs from a reactor-source camera, because it does not use a continuous beam, but is based on 50 pulses/sec of neutrons that are emitted from a uranium or tantalum target after proton bombardment from a synchrotron. In Chapter 5, the ISIS second target station was used for C3 and C3u neutron experiments.

(3.5) Data Interpretation

The data collected through a computer interface at the camera need to be interpreted. The structural information is extracted from Guinier and distance distribution analyses.

(3.5.1) Guinier analyses

Guinier plots are used to analyse experimental data to extract the R_G and R_{XS} parameters and molecular weights. R_G is the radius of gyration which is defined as the root-mean-square distance of all scattering centres in the macromolecule from its centre of gravity. At sufficiently low Q values, the radius of gyration R_G (Figure 3.3 and Figure 3.4) and the forward scattered intensity $I(0)$ are obtained from the slope - $R_G^2/3$ and intercept $ln\ I(0)$ of linear plots :

(Eq. 3.4) $$ln\ I(Q) = ln\ I(0) - R_G^2 Q^2/3.$$

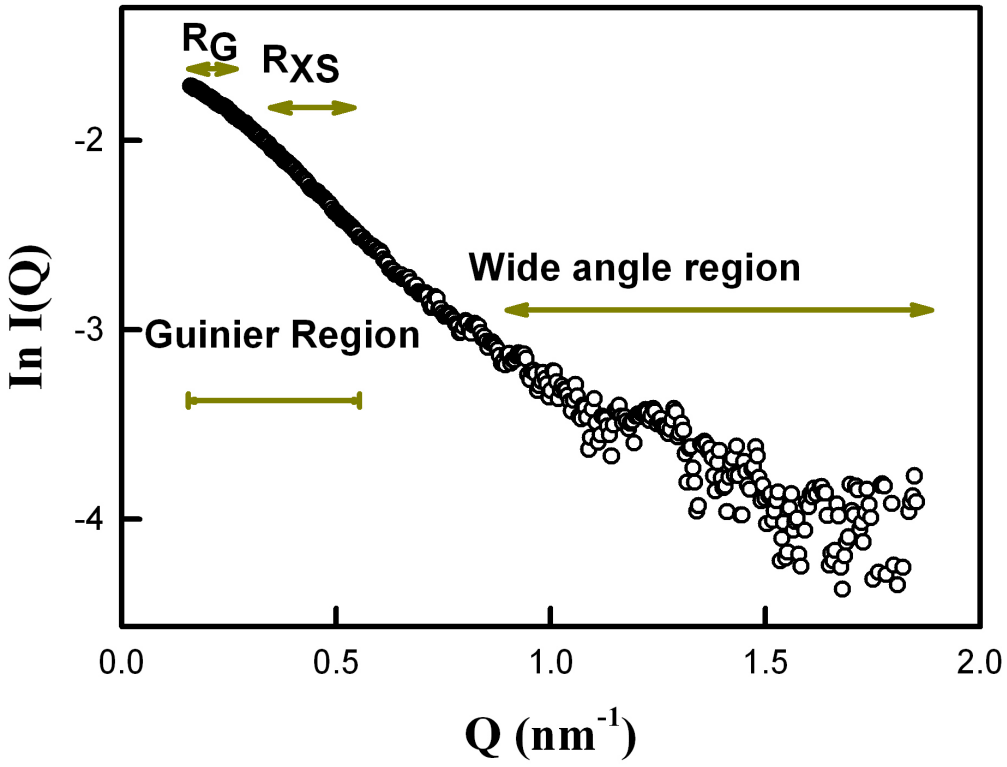


Figure 3.3 General features of a solution scattering curve. In solution scattering curve, the Guinier region gives information for the overall shape of molecule; in the wide angle region secondary and fold (tertiary) structures can be determined. At very low *Q*, the scattering curve is truncated for reason of the beam stop. The scattering curve shown is that obtained for C3u (Chapter 5).

The X-ray R_G value corresponds to a hydrated structure in a highly positive solute-solvent contrast, while the neutron R_G value corresponds to an unhydrated structure and will depend on the solute-solvent contrast. The parameter is analogous to the sedimentation or diffusion coefficients of hydrodynamic experiments to characterise the elongation of the molecules. The smallest Q required for a R_G plot is defined according to $Q.Dmax \leq \pi$, where Dmax is the maximum dimension of the protein. The largest Q required depends on the protein shape, but normally $Q.RG$ is required to be smaller than 1.5.

For an elongated macromolecule, Guinier analyses of the scattering curve at larger Q (Figure 3.3) will give the radius of gyration of the mean cross-section R_{XS} and the cross-sectional intensity at zero angle $[I(Q)Q]_{Q \to 0}$. The Q range is larger the one used for the R_G determination (Figure 3.3):

(Eq. 3.5) $$In\,[I(Q)Q] = In\,[I(Q)Q]_{Q \to 0} - Rxs^2Q^2/2$$

R_{XS} is determined by plotting $In[I(Q)Q]$ against Q^2 to give a straight line of slope $-R_{XS}^2/2$ and intercept $In[(I(Q)Q)]_{Q \to 0}$. Sometimes there are two R_{XS} regions for a protein if this has two distinct types of folded structures. The results of Guinier plots can be validated by analysing the $I(0)$ data. $I(0)/c$ is related with molecular mass of the protein. In X-ray and neutron, high $I(0)/c$ values indicates the occurrence of oligomerisation (Figure 3.4). From $I(0)/c$, we can deduce whether the R_G and R_{XS} increment depends on oligomerisation of the sample or change of the structure. It is important to note that absolute molecular mass values cannot be calculated from X-ray experiments, but can be obtained from neutron studies. Relative mass values are obtained by X-ray scattering.

(3.5.2) R_G and R_{XS} determination of macromolecular dimensions

From R_G and R_{XS}, the maximum length L of a linear model of a protein can be determined (Glatter & Kratky, 1982):

(Eq. 3.6) $$L = \sqrt{12 \cdot (R_G^2 - R_{XS}^2)}$$

Alternatively

(Eq. 3.7) $$L = \frac{\pi I(0)}{[I(Q)Q]_{Q \to 0}}$$

The consistencies of the results are used to validate the presence of rod-like particles.

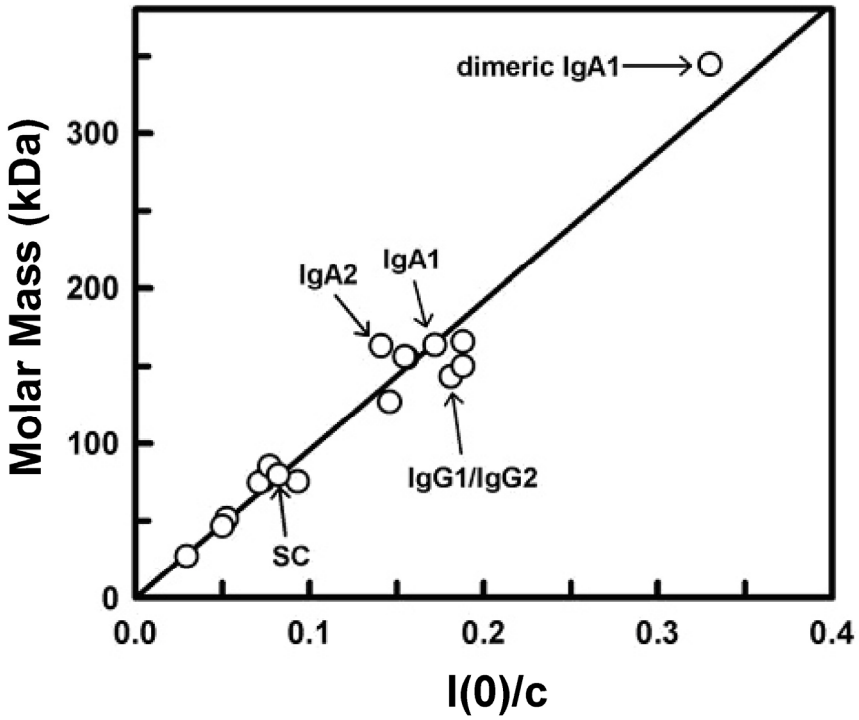


Figure 3.4 Linear relationship between the molar mass and the neutron *I(0)/c* values for glycoproteins in 100% $^{2}H_2O$ buffer (Perkins *et al*., 2008).

(3.5.3) Distance distribution function analyses

The scattering curve $I(Q)$ is a representation in a reciprocal space of the macromolecule. The experimental $I(Q)$ curve is converted to the distance distribution function $P(r)$ by indirect Fourier transformation. $P(r)$ corresponds to the distribution of all the distances r between any two volume elements within the macromolecule. It yields the shape of proteins. This includes the largest particle dimension that is indicated by the r value as the $P(r)$ drops to 0 (Figure 3.5(c,e)) and any notable spatial distributions within the molecule, that are observed as peaks in the $P(r)$ curve. The GNOM program is used as a semi-automated procedure to evaluate *P(r),* and the R_G value from this can be compared with the R_G from the Guinier plot.

In conclusion, solution scattering is important for structural analyses. It overcomes limitations of using other structural determination methods. Scattering does not only provide the solution structures of intact multidomain proteins, but also often provide an opportunity to refine and improve the understanding of crystal structures for proteins. In all these approaches, other complementary structural detection methods can also be used to verify the result of scattering, such as electron microscopy and ultracentrifuge.

(3.6) Biomolecular Modelling

(3.6.1) Introduction to protein structure

(3.6.1.1) Amino acids

Proteins are composed of amino acids which are linked by peptide chain. Every amino acid contains an amino group, a carboxyl group and a side chain (or R group) attached to the central C_α atom. If all four groups are different from each other, the amino acid is asymmetric. For an asymmetrical amino acid, two distinguishable stereoisomers termed L and D forms can co-exist. Only the L-form amino acids are found in proteins. D-form amino acids are present in nature as peptides.

The diversity of the amino acid side chains assigns proteins with a great variety of structures and properties. The amino acids have been grouped into classes according to its side chain (Figure 3.6). Glycine, alanine, isoleucine, leucine, methionine, proline and valine all contain straight, branched arrangements of carbon atoms and are aliphatic side chains. As these side chains become more elongated, there is a greater tendency to be more hydrophobic.

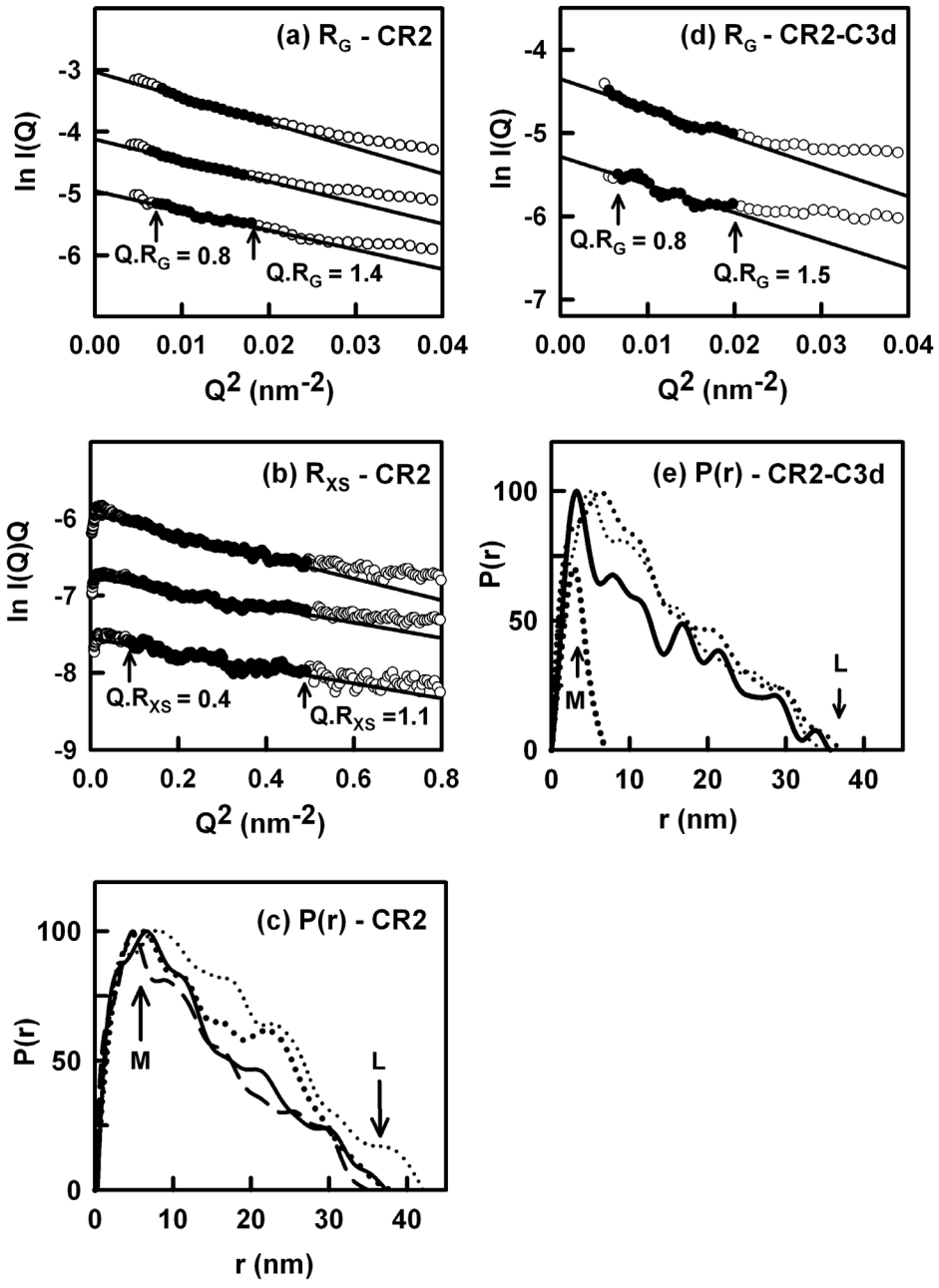


Figure 3.5 Representative X-ray scattering analyses for CR2 and its complex with C3d. The filled circles, in the range enclosed by arrows, represent the regions to obtain R_G and R_{XS} values. (a,d) R_G analyses for CR2 and its complex with C3d. (b) R_{XS} for CR2 only. (c,e) Distance distribution *P(r)* analyses. The maximum in *P(r)* is denoted by *M* and the maximum dimension is denoted by *L* (taken from Chapter 4).

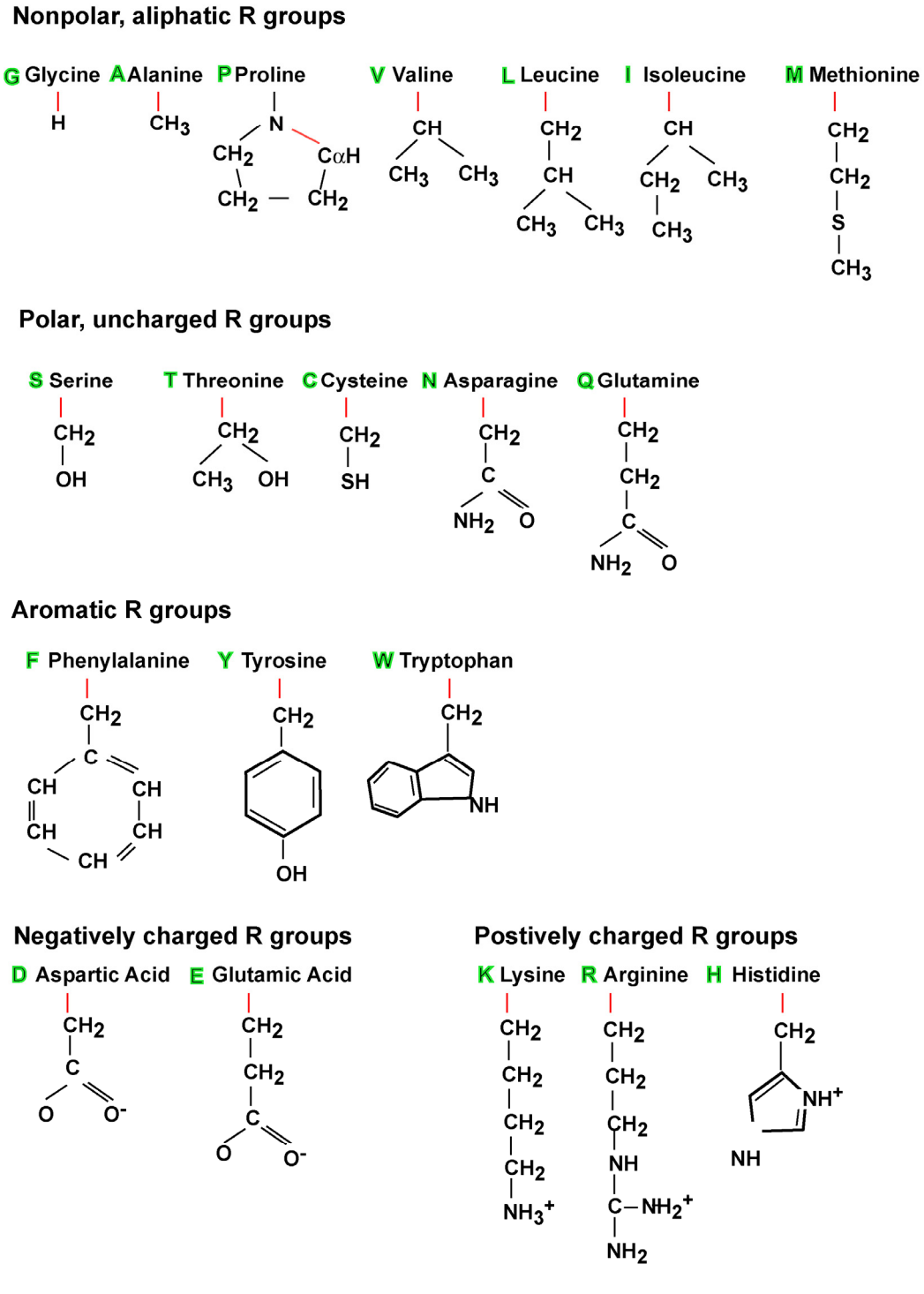


Figure 3.6 The 20 amino acid side chains

The 20 amino acids are divided into groups based on the chemical properties of their side chains (R groups).

Although proline is a cyclic amino acid, it shares many aliphatic properties with other members of this class. Proline can restrict the flexibility of polypeptide chain. On the other side, glycine in this group has no side chain, hence reducing steric collision with other side chains and increasing the flexibility of the polypeptide chain, consequently, glycines frequently appear at turns in the protein structure. Serine, cysteine, threonine, asparagine and glutamine are groups that form hydrogen bonds with water. Serine and threonine have hydroxyl groups, while asparagine and glutamine are the amide derivatives of aspartic and glutamic acids. Cysteine is important for the formation of disulphide bonds. There is also an aromatic group where the members are phenylalanine, tyrosine and tryptophan. Phenylalanine is a hydrophobic amino acid, but tyrosine and tryptophan have polar groups. In addition to the residues mentioned, the remainder have charged side chains (glutamic acid, aspartic acid, histidine, lysine and arginine).

(3.6.1.2) The secondary structure

When a peptide bond is formed, the movement about the C_α-N bond and the C_α-C bond are defined by the phi (Φ) and psi (ψ) angle respectively (Figure 3.7). The conformation of protein backbone can be described by dihedral angles around these bonds for each residue. The combination of phi and psi angles can be displayed as a Ramachandran plot (Figure 3.7b). Most combination of Φ and ψ angles are not allowed because of steric collisions between the side chain and the main chain atoms.

The secondary structures of proteins are α-helices and β-sheets. The α-helices is built from a continuous region of polypeptide chain, whereas the β-sheets is formed from numerous stretches of five to ten residues that are in different sections of the polypeptide chain, but which move into proximity. The allowed Φ and ψ angles for α-helices and β-sheets are described in the Ramachandran plot (Figure 3.7b).

Within protein structure α-helices and β-strands are combined to from a stable hydrophobic core. These compact structure elements are connected by loop regions of various lengths and conformations. The loop regions almost always occur on the surface of the molecule. They rarely form hydrogen bonds with each other; instead they form hydrogen bonds with water molecules. Surveys of known loop structures are not a random collection of possible structures, but a rather limited subset of structure (Sibanda & Thornton, 1985).

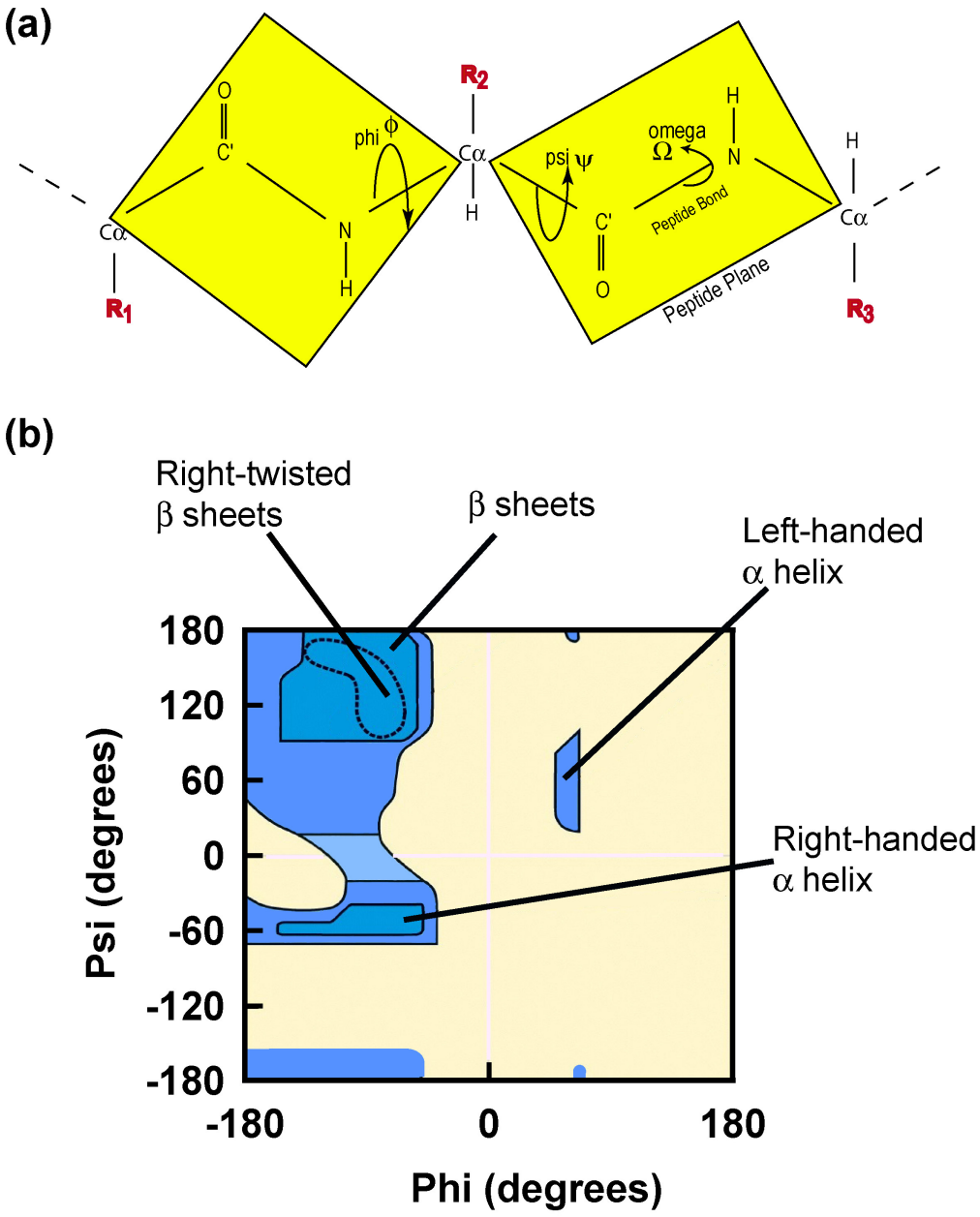


Figure 3.7 Schematic views of the peptide bonds and a Ramachandran plot.

(a) Three residues form two peptide bonds. R represents a side chain of any amino acid residue. The bonds that defines phi, psi and omega (C-N bond) angles are labelled.

(b) Ramachandran plot. The light blue, purple and dark cyan represent area of generously, additional and most favoured regions of C_α-C bond (Drenth, 1999).

A number of combinations of secondary structure elements form with specific geometry arrangements known as super secondary structures. The most common form of super secondary structure is the βαβ motif which involves two parallel β-strands connected by an α-helix. The super secondary structure motifs combine to form compact regions called domains. A domain can be folded into a stable tertiary structure independently of the rest of the protein.

(3.6.2) Constrained scattering modelling

Biomolecular modelling is the creation of models that represent the structure of biological molecules. The fundamental principle of scattering modelling is to confirm that the experimental data has been correctly interpreted. It is useful for multidomain proteins that are not crystallisable (while intact) for reason of interdomain flexibility or heavy surface glycosylation. A unique structure cannot be determined by this method; however, it can identify a group of best-fit conformers that meet well-defined selection criteria (Section 3.6.2.5). In the absence of a crystal structure, the individual domains are modelled using the known structure of homologous domains as templates (Perkins *et al*., 2008). If an atomic structure is available for the domains of the protein, this can be used directly to constrain the analysis. An outline of the modelling procedure is shown in Figure 3.8.

(3.6.2.1) Analysis of protein composition

In order to model the protein, some elements of the protein need to be understood. The program SLUV (Perkins, 1986b) uses the amino acid composition of protein to determine essential protein parameters, such as the protein partial specific volume, molar mass and the molecular volume before and after hydration. It can also calculates the absorption coefficient at a wavelength of 280 nm, which is used for protein concentration estimation (Perkins, 1986b).

(3.6.2.2) Creating atomic models

Multidomain proteins are created by modelling the domains and flexible linkers. If atomic resolution structures for the protein domains are available, then these are used as the basis for the models. If crystal or NMR structures are not available, these are created by homology modelling using a package such as MODELLER. It greatly improves the value of the information provided by structural modelling. For all the models I created, there are

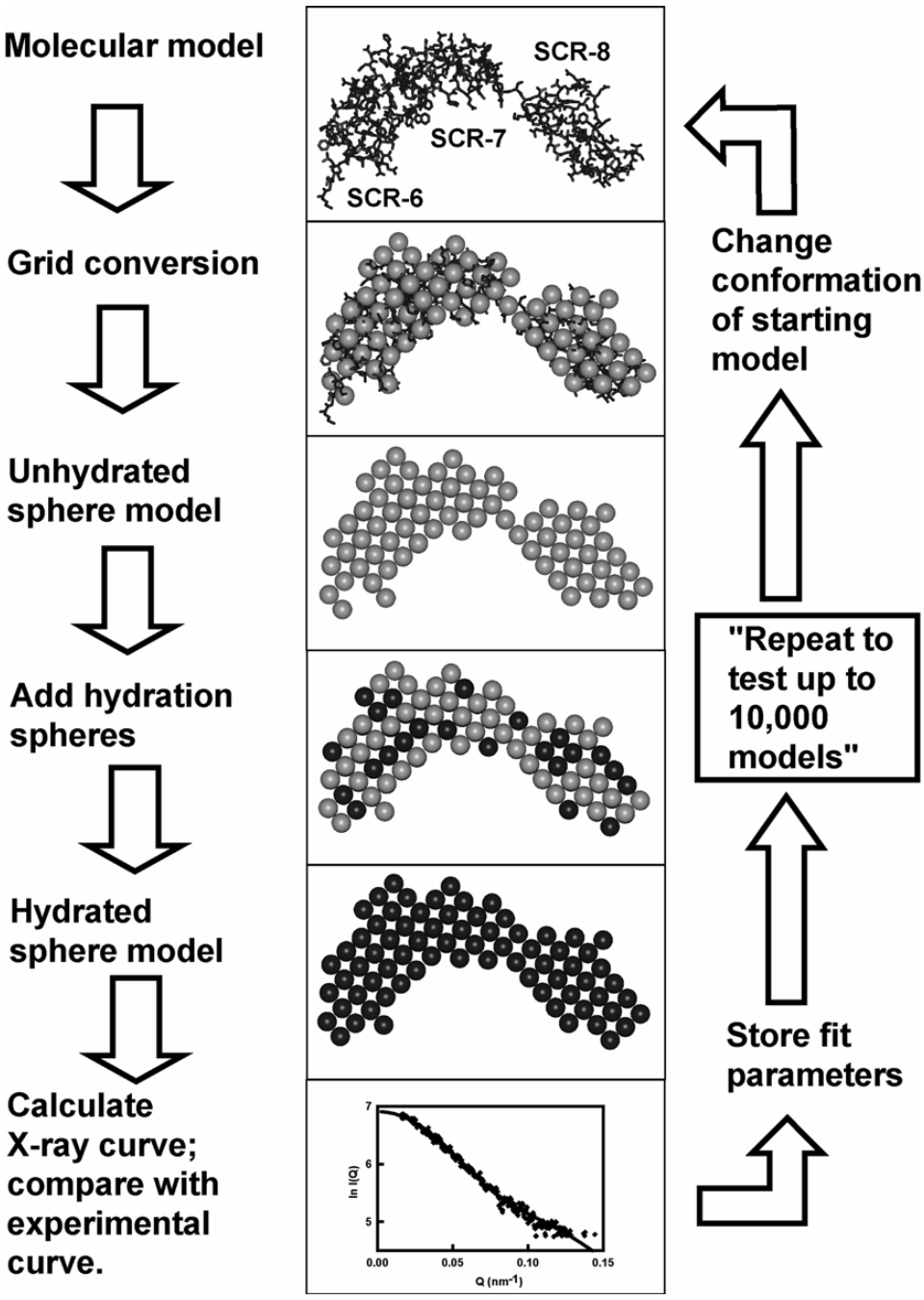


Figure 3.8 Overview of the constrained modelling procedure (Perkins *et al*., 2008).

crystallised domains as templates.

The models were created in randomised arrangements to present all possible orientation of domains within the protein. This is achieved by a library of different conformational linkers, which are the peptides outside the fixed domain structure. A library of thousands of random linker conformations were generated by INSIGHT II 98.0 molecular graphics software (Accelrys, San Diego, CA, USA). It is generated by fixing the temperature of 298K and the linker length, which is measured by the distance between the first and last Cα atom, is either allowed to vary at will or restricted to certain values. In the automated modelling procedure, linkers are then randomly selected from the linker library to join the domains together to generate a series of about 2000-8000 intact models (Figure 3.8). Usually, even amount of linkers with diverse lengths are the best for creating randomised models.

(3.6.2.3) Debye sphere modelling

The next stage is to convert the 2000-5000 three-dimensional models into sphere models consisting of small non-overlapping spheres of uniform density and with the same total sphere volume as the original atomic model (Figure 3.8, Perkins *et al*., 2002). If the atomic coordinate model is available, the coordinates are converted directly to spheres by placing all atoms within a three dimensional grid of cubes of side around 0.6 nm. The size of the cubes and the number of atoms that each cube contains (namely cut-off level) are defined and tested by trial and error. The grid transformation of the resulting sphere model needs to have a similar volume as the unhydrated volume of the protein, which is calculated from the protein composition (including carbohydrates) by the program SLUV (Section 3.6.2.1). If residues are missing in the atomic model, the volume discrepancy can be compensated by adjustment of the sphere cut-off during sphere generation. For the hydrated protein model, a layer of additional spheres is evenly added onto the surface of the dry sphere model, where the total number of spheres gives the hydrated volume (Perkins, 1986b; Ashton *et al*., 1997; Perkins *et al*., 2008). Once the sphere size has been determined, it is kept constant for the automated curve fit searches.

(3.6.2.4) Debye scattering curve calculation

Curves are calculated from the sphere models by Debye's Law (Eq. 3.2) adapted to

small spheres (Perkins et al., 1998):

(Eq. 3.8) $$\frac{I(Q)}{I(0)} = g(Q)\left(n^{-1} + 2n^{-2}\sum_{j=1}^{m} A_j \frac{\sin Qr_j}{Qr_j}\right)$$

(Eq. 3.9) $$g(Q) = (3(\sin QR - QR\cos QR))^2 / Q^6R^6,$$

Where $g(Q)$ is the squared form factor for the sphere of radius R, n is the number of spheres filling the body, A_j is the number of distances r_j for that value of j, r_j is the distance between the spheres, and m is the number of different distances r_j. The equation is modified from Debye's equation (Eq. 3.2) to consider the spheres of the model rather than the electron or individual scattering from each nucleus in the molecule. These spheres have to be sufficiently small (about 0.6 nm) so that their form factor in the Debye equation is almost invariant in the Q scattering range used. The curves generated from this calculation can be compared directly with experimental curves to determine best-fit models.

(3.6.2.5) Identification of best-fit models

The identification of the best-fit models is achieved by the use of four filters:

(a) Remove the models which have significant steric overlaps between domains or subunit, because of the randomisation procedure. Steric overlap means that the number of spheres is much reduced compared to that predicted from its composition. Normally any models which have spheres less than 95% of the volume calculated from the composition needs to be removed.

(b) The R_G and R_{XS} values calculated from each model have to agree with experimental Guinier fits in the same Q ranges used (values within 5% of the experimental values are good models).

(c) An R factor is calculated for a quantitative comparison of each model scattering curve $I(Q)_{cal}$ against experimental scattering data $I(Q)_{exp}$ over the whole Q range, where $I(Q)_{cal}$ is set as 1000:

(Eq. 3.10) $$R = \frac{\Sigma \mid I(Q)_{exp} - I(Q)_{cal} \mid}{\Sigma \mid I(Q)_{exp} \mid} \times 100\%$$

The R factor value reflects the goodness-of-fit between the calculated and experimental curves. It should be less than 10% for a good fit. With these criteria, the best family of structures can be narrowed down.

(d) Sedimentation coefficients of the models can be calculated from the hydrated sphere models. They were used to compare with the experimental values as an independent monitor.

(3.6.2.6) Calculation of hydrodynamic properties

The sedimentation coefficient of all the models can be calculated from the hydrated scattering sphere models using the program HYDRO (Garcia de la Torre *et al*., 1994) and also directly from the atomic coordinates in the HYDROPRO using the default value of 0.31 nm for the atomic element radius for all atoms that includes the hydration shell (Garcia de la Torre *et al*., 2000). Comparison of the sedimentation coefficients of the models with that experimentally obtained by SV experiments provides another way to validate the scattering modelling.

Chapter Four

Solution structure of the complex formed between human complement C3d and full length complement receptor Type 2

(4.1) Introduction

Complement receptor type 2 (CR2/CD21), a type I membrane glycoprotein found on cell surfaces, belongs to the regulators of complement activation protein family (Law & Reid, 1995; Morgan & Harris, 1999). CR2 is comprised of 15 or 16 SCR domains depending on alternative splicing (Toothaker *et al.,* 1989; Fujisaku *et al.,* 1989), a transmembrane region and a short cytoplasmic tail. Cells that express CR2 include mature B-cells, follicular dendritic cells, epithelial cells, and some T cells. CR2 has four known ligands: (i) the C3d (Figure 4.1) and iC3b cleavage fragments of C3 (Lowell *et al.,* 1989; Carel *et al.,* 1990); (ii) the coat glycoprotein gp350 of the Epstein-Barr virus (Figure 4.1; Fingeroth *et al.,* 1984); (iii) interferon alpha IFNα (Figure 4.1; Delcayre *et al.,* 1991; Asokan *et al.,* 2006); and (iv) CD23, a low affinity receptor for IgE (Figure 4.1; Aubry *et al.,* 1992). The complement C3d fragment is a 35 kDa cleavage product of the central complement protein C3, which is formed by the regulatory breakdown of C3b by factor I and factor H; it is derived from the thioester-containing domain of C3 (Janssen *et al.,* 2005). CR2 is important for B-cell activation and the generation of the immune response by linking the innate and adaptive immune systems. The prevailing model of CR2 function reveals that binding of C3d to CR2 on mature B lymphocytes together with the involvement of CR2 in a complex with CD19, CD81 and mIgM reduces the threshold of immune activation. Related activities include intracellular reactions, including an increase in intracellular Ca^{2+} (Carter *et al.,* 1988, 1991), the activation of MAP kinases (Tooze *et al.,* 1997), and B-cell proliferation (Carter & Fearon, 1992).

Diverse structural approaches have clarified CR2 function. Hydrodynamic and electron microscopy data on recombinant CR2 showed extended conformations (Moore *et al*., 1989). The binding sites for its four ligands were located to both the SCR 1-2 domains (and in the case of CD23 to the SCR 5-8 domains as well) (Ahearn *et al.,* 1988; Carel *et al*., 1990; Martin *et al.,* 1991; Aubry *et al.,* 1994). It is of great interest is that the compact V-shaped arrangement of the SCR1-2 domains of CR2 in its complex with C3d seen in its crystal structure suggested that only the SCR 2 domain interacted with C3d (Szakonyi *et al*., 2001). The unbound CR2 SCR 1-2 domains also showed a similar compact V-shaped crystal structure (Prota *et al*., 2002). In distinction to this, three solution structures for unbound CR2 SCR 1-2 in 50 mM NaCl revealed an open V-shaped structure (Guthridge *et al*., 2001; Gilbert *et al*., 2005, 2006a). Unlike the co-crystal structure, studies of the CR2 SCR 1-2 complex with C3d by constrained

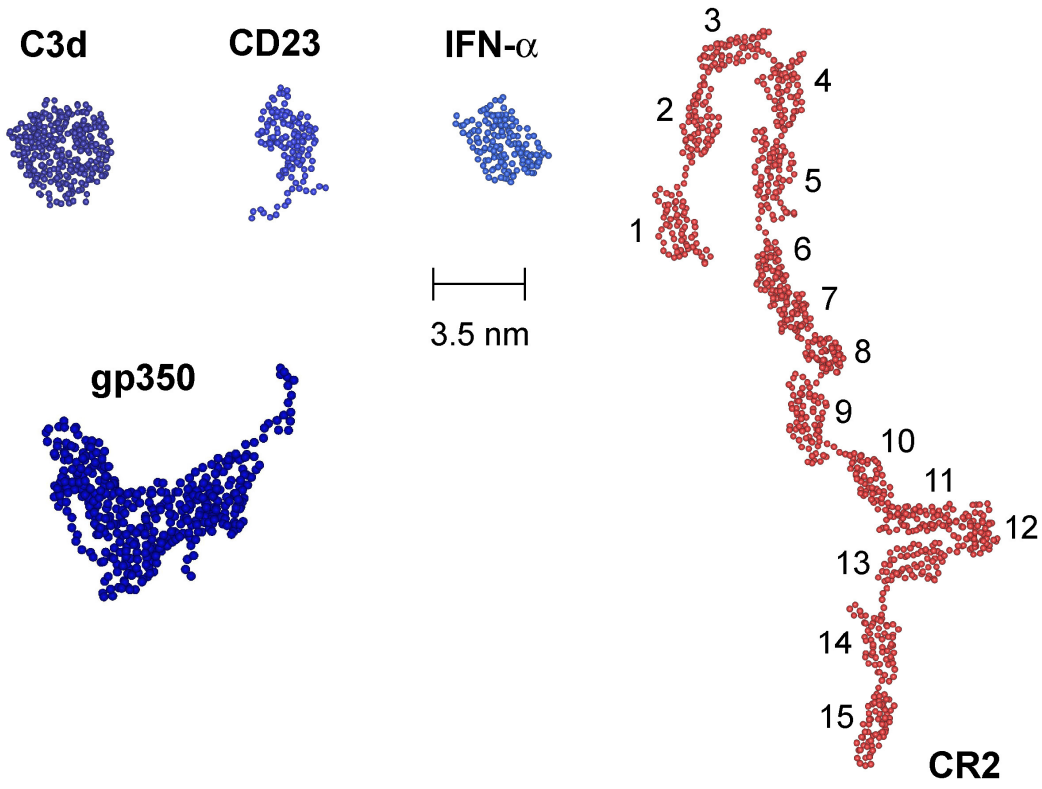


Figure 4.1 Views of CR2 SCR 1-15 in comparison with its ligands, namely C3d, the C-type lectin domain of CD23, IFN-α and gp350. The SCR domains are numbered from 1 to 15 in the CR2 model (red, blue in Figure 4.6). The ligands (blue) are based on crystal and NMR structures with the following PDB codes 1GHQ, 1T8D, 1RH2 and 2H6O (Gilbert *et al*., 2006b).

scattering modelling in 50 mM NaCl and mutagenesis in 137 mM NaCl, showed that both the SCR 1 and SCR 2 domains were bound to the surface of C3d (Gilbert *et al*., 2005; Hannan *et al*., 2005). Hence, in comparison, both crystal structures either potentially demonstrate artefactual arrangements of SCR domains, or indicate that SCR 1-2 exists in alternative structural conformations depending on its biological context. The solution structural determination of full length CR2 SCR 1-15 in 137 mM NaCl revealed an extended and flexible structure with some regions of folding back, and this flexibility is likely to facilitate a broad range of structural interactions between C3d-antigen complexes and the B-cell receptor (Figure 4.1; Gilbert *et al*., 2006b).

The interaction of CR2 SCR 1-2 with C3d demonstrates a substantial dependence on charge. Surface plasmon resonance studies in 50 mM NaCl that included SCR 1-2, SCR 1-4 and SCR 1-15 demonstrated a high affinity for C3d, but in more physiological 125 mM NaCl salt, the binding between CR2 and C3d is much weakened (Asokan *et al.,* 2006; Moore *et al.,* 1989; Guthridge *et al.,* 2001; Hannan *et al.,* 2005). Many earlier binding studies with CR2 and C3d had employed 50 mM NaCl buffers, and a broad range of dissociation constants K_D had been determined (Moore *et al.,* 1989; Guthridge *et al.,* 2001; Hannan *et al.,* 2005; Asokan *et al.,* 2006). Some anomalous binding effects had been observed (Asokan *et al.,* 2006; Sarrias *et al*., 2001).

Knowledge of the solution structure of the CR2-C3d complex is important for understanding the role of the 15 SCR domains in CR2 function. It is also important to know how the formation and structure of the CR2-C3d complex is modified by electrostatic effects. Here, the joint use of analytical ultracentrifugation (AUC) and X-ray solution scattering as independent low resolution structural tools suggests that CR2 SCR 1-15 (denoted as CR2 below unless specified otherwise) is almost unaltered in overall conformation between 50 mM and 137 mM NaCl. Unexpectedly, we show that the presumed dimerisation of C3d reported in earlier work (Gilbert *et al.,* 2005) is in fact an equilibrium between monomers, dimers and trimers in 50 mM NaCl, and that C3d becomes monomeric in 137 mM NaCl. The use of both 50 mM and 137 mM NaCl buffers to study complex formation between full-length CR2 and C3d shows that the CR2 conformation is not affected by C3d binding, but that complex formation is detectable only in 50 mM NaCl. This places limits on previously-determined K_D values. We discuss the implications of these results for biochemical affinity studies and the physiological functioning of CR2.

(4.2) Results and Discussion

(4.2.1) Sedimentation velocity of unbound CR2

Analytical ultracentrifugation studies macromolecular structures in solution by following their sedimentation behaviour on subjecting these to a high centrifugal force (Cole *et al.*, 2008). By sedimentation velocity, the sedimentation coefficient $s^0_{20,w}$ monitors macromolecular elongation and sample polydispersity. Immediately before ultracentrifugation, CR2 was subjected to gel filtration on a Superdex-200 column to remove pre-existing aggregates or degradation products. All CR2 samples were checked before and after ultracentrifugation experiments by SDS-PAGE.

In order to evaluate the effect of NaCl concentration on CR2, sedimentation velocity runs were performed in both PBS buffer (137 mM NaCl) and HEPES buffer (50 mM NaCl) at three concentrations between 0.12-0.34 mg/ml and four rotor speeds. The different rotor speeds established reproducibility and any speed-dependent flexibility in the multidomain CR2 structure. Lamm equation fits of up to 200 scans using the continuous *c(s)* distribution model in SEDFIT (Section 4.4.2) resulted in good visual agreements between the experimental and fitted CR2 boundaries and satisfactory low root-mean-square deviations. A single species was observed by both interference and absorption optics to confirm its monomeric structure. The *c(s)* distributions in Figure 4.2 showed that the CR2 $s^0_{20,w}$ value was similar within error at 4.0 S in 137 mM NaCl (Figure 4.2(a,c)) and 4.1 S in 50 mM NaCl (Figure 4.2(b,d)). The mean of all the $s^0_{20,w}$ values in 137 mM NaCl was 3.93 ± 0.20 S (11 determinations) and in 50 mM NaCl was 4.03 ± 0.13 S (4 determinations). The present determinations are in good agreement with the previous studies of CR2 in 137 mM NaCl that gave 4.24 ± 0.1 S using DCDT+ software (Gilbert *et al.*, 2006b). The similar sedimentation coefficients show that the multidomain conformation of CR2 was not altered by ionic strength. If the slight reduction of 0.1 S on passing from 50 mM NaCl to 137 mM NaCl is significant, this would correspond to a minor lengthening of the CR2 multidomain structure in 137 mM NaCl.

(4.2.2) Sedimentation velocity of unbound C3d

Recombinant C3d was purified by thrombin cleavage of a C3d-GST construct (Section 4.4.2). Previously, C3d purified from a pET15b expression vector was observed to undergo what was assumed to be a concentration-dependent monomer-dimer equilibrium with a K_D value of 40 μM in 10 mM HEPES, 50 mM NaCl (Gilbert *et al.*, 2005). Here,

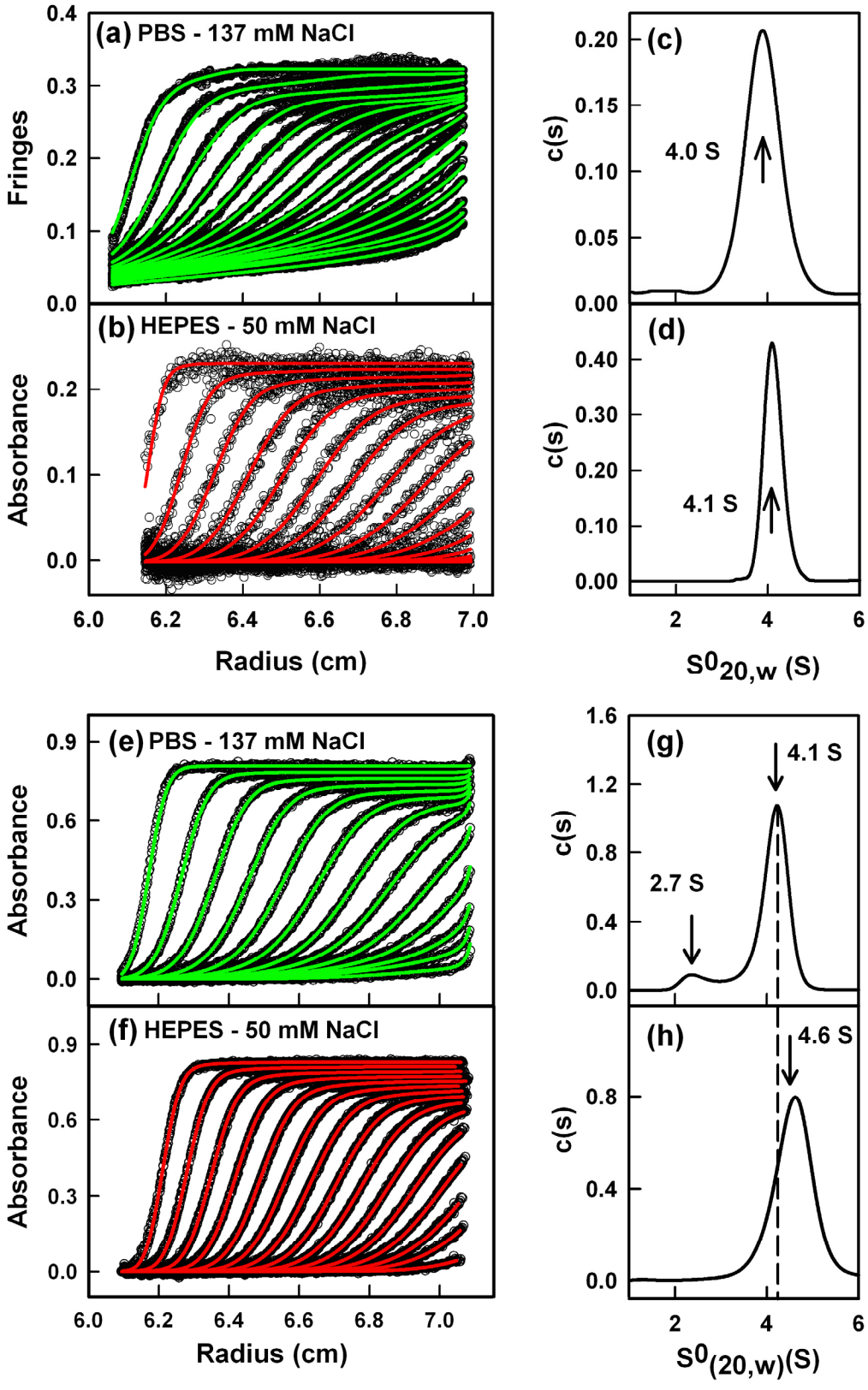


Figure 4.2 (Legend overleaf)

Figure 4.2 Sedimentation velocity data analyses for unbound CR2 and its 1:1 complex with C3d in PBS (green) and HEPES (red) buffer by Lamm equation fits using SEDFIT.

(a, c) The boundary fits for CR2 SCR1-15 at 0.12 mg/ml in 137 mM NaCl at 30,000 r.p.m. are from interference data optics. Only every sixth scan of the first 100 scans is shown for clarity. The *c(s)* distribution plot showed one sedimenting species with a $s^0_{20,w}$ value of 4.0 S for 137 mM NaCl.

(b, d) The boundary fits for CR2 SCR1-15 at 0.15 mg/ml in 50 mM NaCl at 30,000 r.p.m. are from 280 nm absorbance data optics. Only every seventh scan is shown for clarity. The *c(s)* distribution plot showed one sedimenting species with a $s^0_{20,w}$ value of 4.1 S for 50 mM NaCl.

(e, g) The boundary fit (green) for the complex at 0.72 mg/ml in 137 mM NaCl gave two peaks at $s^0_{20,w}$ values of 2.7 S and 4.1 S as shown. Only every fifth scan of the first 65 scans at 42,000 r.p.m. is shown for clarity.

(b, d) The boundary fit (red) for the complex at 0.46 mg/ml in 50 mM NaCl gave only one peak at a $s^0_{20,w}$ value of 4.6 S. Only every fifth scan of the first 75 scans in 50 mM NaCl at 42,000 r.p.m. is shown for clarity.

this was reinvestigated by sedimentation velocity at concentrations from 0.3 mg/ml to 8.3 mg/ml in both 137 mM and 50 mM NaCl buffers at three different rotor speeds (Section 4.4.2). Interference optics data were used throughout, together with absorbance data below 0.8 mg/ml, due to the saturation of the absorbance optics at C3d concentrations above 0.8 mg/ml. Unlike the previous DCDT+ analyses, size-distribution *c*(*s*) analyses using SEDFIT were now used for data analyses, and resulted in good boundary fits (Figure 4.3(a,b)). In 137 mM NaCl, a single predominant species was observed at an $s^0_{20,w}$ value of 3.0 ± 0.1 S between 0.5-7.5 mg/ml (ten determinations) (Figure 4.3c). The *c*(*M*) size distribution plot showed that this corresponded to a molar mass of 39 ± 4 kDa. This was accordingly assigned to C3d monomers with a molecular mass of 34.6 kDa calculated from its composition. Trace amounts of higher oligomers were visible at higher C3d concentrations (Figure 4.3c). In 50 mM NaCl, the *c*(*s*) distributions showed a concentration dependence between 0.3mg/ml and 8.3 mg/ml, and three peaks indicating three oligomer species were visible (Figure 4.3d). At 0.3 mg/ml, only the monomer (M) was present at 2.9 S. The mean value is 3.0 ± 0.3 S (21 determinations). Above 0.7 mg/ml, a second peak (D) was observed at 4.3 ± 0.3 S (16 determinations) and its intensity increased with concentration. A third peak (T) appeared above 2.1 mg/ml, whose $s^0_{20,w}$ value increased from 5.4 S to 6.8 S as the concentration increased from 2.1 mg/ml to 8.3 mg/ml. The mean value is 6.0 ± 0.6 S (11 determinations). The *c*(*M*) distribution plots showed that peaks M, D and T correspond to molecular masses of 34 ± 4 kDa, 59 ± 10 kDa and 95 ± 18 kDa respectively, assuming that the frictional ratio *f/fo* remained unchanged for these species (Table 4.1). These masses are consistent with the formation of C3d dimers and trimers of sizes 70 kDa and 105 kDa. The rate of interconversion between the three species was slow enough during the velocity runs for these peaks to be visible in the *c(s)* plots. Our previous study of C3d in 50 mM NaCl buffer using DCDT+ fits of 10-20 boundaries gave $s^0_{20,w}$ values that ranged between 3.26 S to 4.85 S. (Gilbert *et al*., 2005). This range of $s^0_{20,w}$ values is consistent with the range seen in Figure 4.3, and reanalyses of the previous data using SEDFIT confirmed the occurrence of three species in Figure 4.3d.

Notably, in Figure 4.3d, at 8.3 mg/ml; the third peak which is labelled as "T" may be a mixture of trimer and tetramer. C3d is a globular protein. It is known for globular proteins that the sedimentation coefficient ratios between monomer-dimer, monomer-trimer and monomer-tetramer can be fitted by the rule of: (1) The sedimentation coefficient of the dimer is $4^{1/3}$ times that of the monomer. (2) The

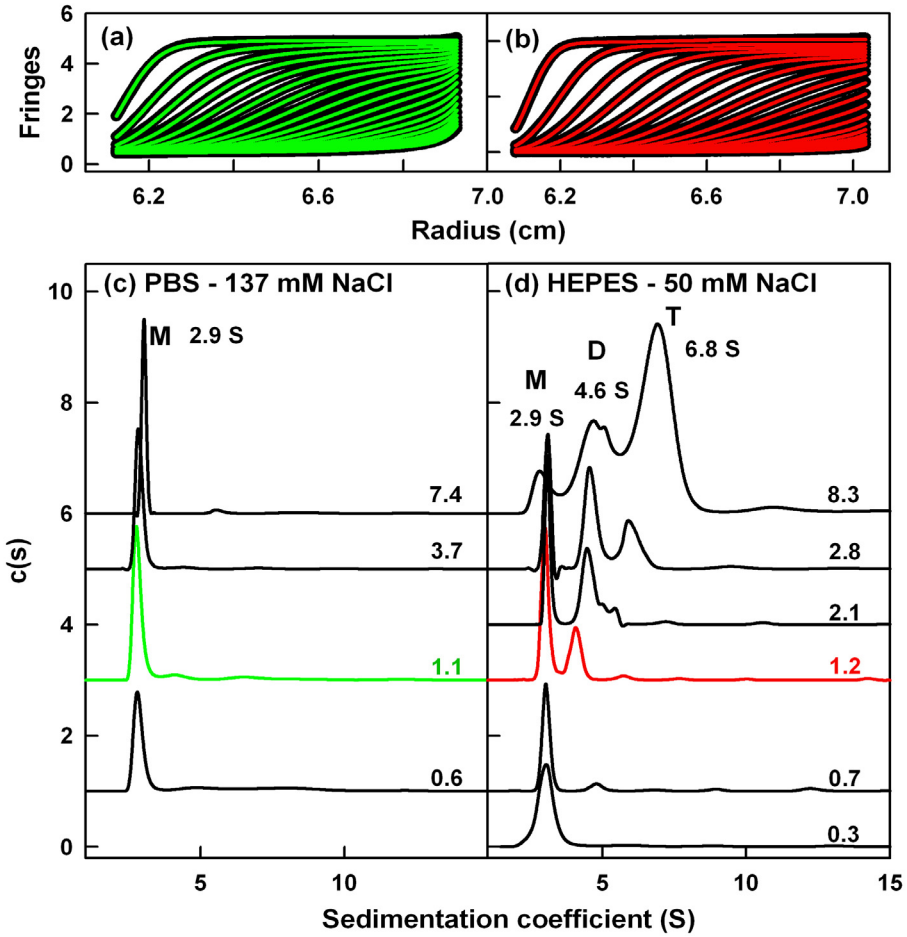


Figure 4.3 Concentration dependence of the sedimentation velocity data for free C3d in PBS (green) and HEPES (red) buffers. Figure **(b)** and **(d)** are used with permission from Dr A. I. Okemefuna (Okemefuna, 2009).

(a) The boundary fits for C3d at 1.1 mg/ml in 137 mM NaCl at 32,000 r.p.m. are from interference data. Only every sixth scan in a total of 200 scans is shown for clarity.

(b) The boundary fits for C3d at 1.2 mg/ml in 50 mM NaCl at 40,000 r.p.m. are from interference data. Only every fourth scan of the first 80 scans is shown for clarity.

(c) The concentration dependence of the *c*(*s*) size distributions for 0.6 mg/ml, 1.1 mg/ml (green), 3.6 mg/ml and 7.4 mg/ml of C3d in 137 mM NaCl is shown. A single species is observed at 2.9 S and labelled M (monomer).

(d) The concentration dependence of the *c*(*s*) size distributions for 0.3 mg/ml, 0.7 mg/ml, 0.8 mg/ml, 1.2 mg/ml (red), 2.1 mg/ml, 2.8 mg/ml and 8.3 mg/ml of C3d in 50 mM NaCl is shown. The peaks are labelled as M (monomer), D (dimer) and T (possible trimer) in increasing order of S values, where the values shown correspond to 8.3 mg/ml.

Table 4.1 Sedimentation coefficients for CR2, C3d and their complex.

Protein	Buffer	Experimental $s^0_{20,w}$ (S)	Frictional ratio *f/fo*	Calculated $s^0_{20,w}$ (S)
C3d monomer	10 mM HEPES, 50 mM NaCl, pH 7.4	3.0 ± 0.3	1.066	3.1, 2.9 **
C3d monomer	9.6 mM phosphate, 137 mM NaCl, pH 7.4	3.0 ± 0.1		
C3d dimer	10 mM HEPES, 50 mM NaCl, pH 7.4	4.3 ± 0.3	1.181	4.7, 4.5 **
C3d trimer	10 mM HEPES, 50 mM NaCl, pH 7.4	6.0 ± 0.6	1.106	5.9, 5.8 **
CR2	10 mM HEPES, 50 mM NaCl, pH 7.4	4.03 ± 0.13	1.99 ± 0.15*	3.94
CR2	9.6 mM phosphate, 137 mM NaCl, pH 7.4	3.93 ± 0.20	2.06 ± 0.27*	3.89
CR2-C3d complex	10 mM HEPES, 50 mM NaCl, pH 7.4	4.52 ± 0.11	2.11 ± 0.27*	4.57 ± 0.04
CR2-C3d complex	9.6 mM phosphate, 137 mM NaCl, pH 7.4	2.52 ± 0.40, 4.07 ± 0.14	n.a.	n.a.

* The frictional ratio was floated in the SEDFIT analyses for CR2 and the CR2-C3d complex.

** The first value is from HYDRO; the second is from HYDROPRO

n.a. – not available

sedimentation coefficient of the trimer is $9^{1/3}$ times that of the monomer. (3) The sedimentation coefficient of the tetramer is $16^{1/3}$ times that of the monomer (Cölfen & Völkel, 2003). If the C3d monomer has a sedimentation coefficient of 2.9 S, the predicted dimer, trimer and tetramer sedimentation coefficients will be 4.6 S, 6.03 S and 7.31 S respectively. In 50 mM NaCl buffer (Figure 4.3d), the observed dimer value of 4.6 S for C3d agrees with the calculated value of 4.6 S. At 8.3 mg/ml, the observed *s* value of 6.8 S is intermediate between that calculated for the trimer of 6.03 S and that for the tetramer of 7.31 S. The *c*(*M*) distribution plots showed that peak T correspond to a molecular mass of 95 ± 18 kDa, which is close to that of 102 kDa for a trimer; therefore, we cannot state if C3d above 2.8 mg/ml is a tetramer or trimer only or a mixture of tetramer and trimer.

(4.2.3) Sedimentation equilibrium of unbound C3d

Sedimentation equilibrium experiments in both 137 mM NaCl and 50 mM NaCl were performed to explore C3d oligomer formation over a ten-fold concentration range. Absorbance data were not fitted because of the saturation of the optics above 0.8 mg/ml. The individual fits of the equilibrium curves at different speeds and concentrations gave the molecular weights shown in Figure 4.4c. Thus C3d in 50 mM NaCl showed a concentration dependent mass, while this was not seen in 137 mM NaCl buffer. The final global fit of each multispeed experiment was done at rotor speeds of 19,000, 28,000 and 35,000 r.p.m. (Section 4.4.2).

For C3d in 137 mM NaCl, the mass was determined using a global fit of 15 curves at five concentrations (Figure 4.4a) on the basis of a monomer-dimer-trimer model (Figure 4.3). Low random residuals, a low reduced χ^2 value of 8.0, and good visual fits were obtained. A C3d monomer mass of 34.1 ± 1.6 kDa was determined, in good agreement with the composition-determined value of 34.6 kDa. The monomer-dimer and monomer-trimer dissociation constants K_D were high at 61 mM and 1.4 mM^2 respectively, indicating that no dimer or trimer had formed. This outcome was confirmed by a global fit based on a single species only, which gave a similar mass of 35.1 ± 0.3 kDa and an almost unchanged reduced χ^2 value of 9.7 (fits not shown). Hence, like CR2, C3d was concluded to be monomeric in physiological buffer.

For C3d in 50 mM NaCl buffer, the same rotor speeds and concentration range also gave good fits to a monomer-dimer-trimer association model. This fit required that

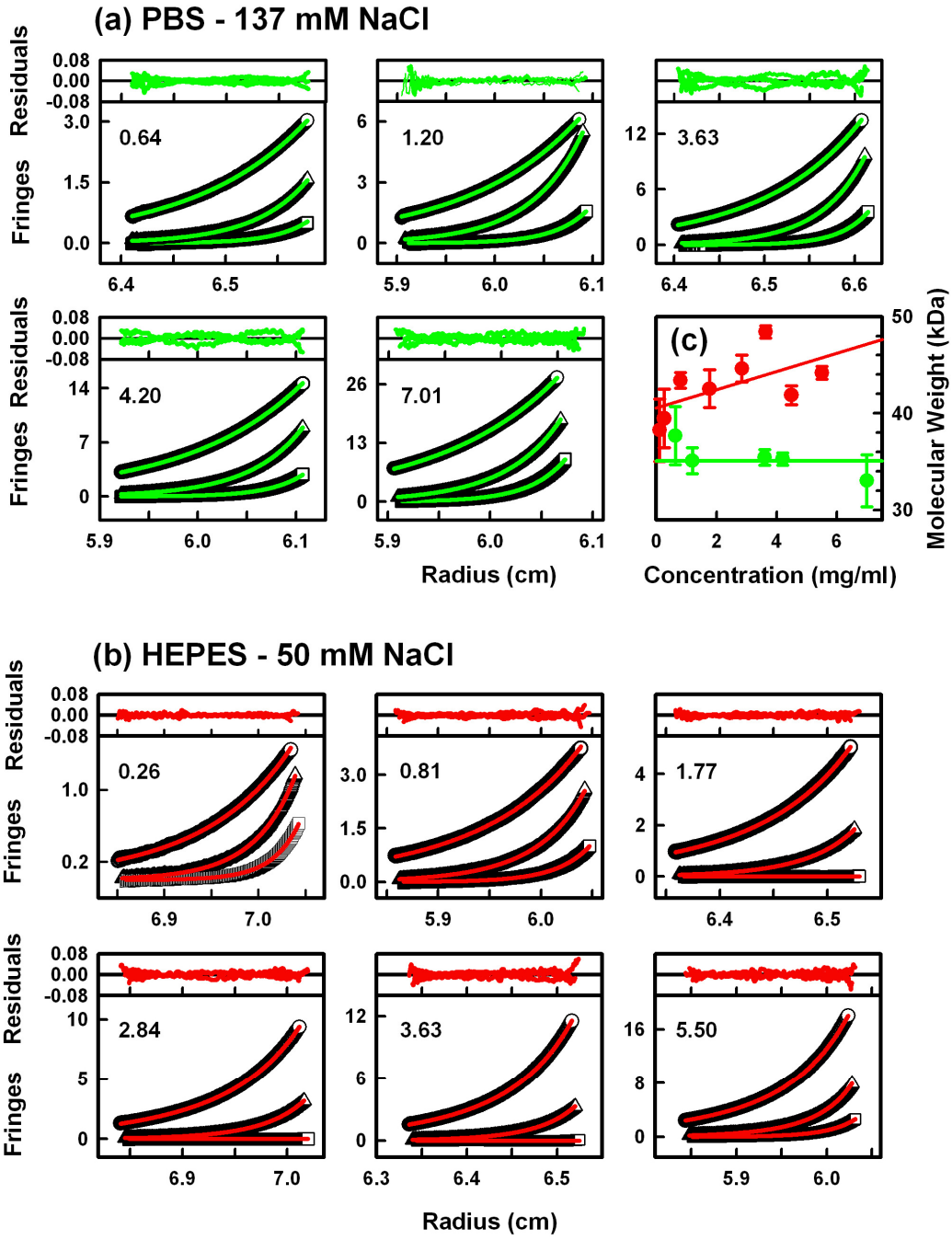


Figure 4.4 (Legend overleaf).

Figure 4.4 Multispeed global fit sedimentation equilibrium analyses of C3d in 137 mM NaCl (green) and 50 mM NaCl (red). The fit residuals are shown above each panel. Figure **(b)** is used with permission from Dr A. I. Okemefuna (Okemefuna, 2009).

(a) The five panels show the multispeed global fit analysis of interference data for C3d (fitted line in green) at five concentrations (0.64 mg/ml, 1.20 mg/ml, 3.63 mg/ml, 4.20 mg/ml and 7.01 mg/ml). Within each panel, the rotor speeds were 19k r.p.m. (○), 28k r.p.m (Δ) and 35k r.p.m (□). The global fits gave a monomer mass of 34.1 kDa.

(b) The six panels show the multispeed global fit analysis of interference data for C3d (fitted line in red) at six concentrations (0.26 mg/ml, 0.81 mg/ml, 1.77 mg/ml, 2.84 mg/ml, 3.63 mg/ml and 5.50 mg/ml). The rotor speeds are the same as in (a).

(c) Concentration dependence of the molecular mass of C3d in 137 mM (green) and 50 mM (red) NaCl buffers. For 137 mM NaCl, the green line corresponds to a mass of 35 kDa. For 50 mM NaCl, the red line corresponds to the regression line drawn using data from 0.11 mg/ml, 0.26 mg/ml, 0.81 mg/ml, 1.77 mg/ml, 2.84 mg/ml, 3.63 mg/ml, 4.49 mg/ml and 5.5 mg/ml.

equilibrium had been reached, and this was confirmed using overlays of runs (Section 4.4.3). The global fit based on a monomer-dimer-trimer association model of 18 curves at six concentrations with a fixed monomer mass of 34.6 kDa gave low residuals (Figure 4.4b) and a global reduced χ^2 value of 3.5. The K_D values of the monomer-dimer and monomer-trimer equilibria were determined to be 360 ± 0.01 μM and 0.070 ± 0.003 μM^2 respectively. Global fits based on a single species gave much worsened reduced χ^2 values of 60 to 223. A monomer-dimer model gave a poorer reduced χ^2 value of 12.9 but gave a K_D value of 210 μM, which was comparable with the previous estimate of 40 μM (Gilbert *et al*., 2005). Overall, C3d at 8.3 mg/ml in low salt will form more trimer than dimer, and this agrees with the *c(s)* analyses of the sedimentation velocity data (Figure 4.3d).

(4.2.4) Sedimentation velocity of the CR2-C3d complex

Previous studies by surface plasmon resonance and mutagenesis studies showed that the CR2-C3d complex is formed with a K_D of 22-27 nM in 50 mM NaCl, and that this association is weakened as [NaCl] is increased (Guthridge *et al*., 2001; Hannan *et al*., 2005). Accordingly the CR2-C3d complex was studied by sedimentation velocity in both 50 mM NaCl and 137 mM NaCl using a 1:1 molar ratio of CR2 and C3d.

In 50 mM NaCl, the boundary fits for the complex in *c*(*s*) plots in a dilution series starting from 0.54 mg/ml (a range of 0.28-3.5 μM) showed a single peak at a $s^0_{20,w}$ value of 4.52 ± 0.11 S (4 determinations) (Figure 4.2(f,h)). As this peak has shifted from a value of 4.0-4.1 S for unbound CR2 (Figure 4.2(c,d)), this single peak indicated that full complex formation in the fluid phase had occurred between 0.28 μM and 3.5 μM, and therefore that its K_D value is below 0.28 μM. The *c*(*M*) plot from Figure 4.2(f,h) indicated a molecular mass of 140 ± 17 kDa, which is close to the expected value of 155 kDa calculated from its composition. The fits showed that the frictional ratios of CR2 and the CR2-C3d complex were similar (Table 4.1). This indicates that no conformational change in CR2 had occurred with complex formation, and that the peak shift arises as the result of the change in molecular mass.

In 137 mM NaCl, the corresponding *c*(*s*) fit for the complex in a dilution series starting from 1.08 mg/ml (a range of 0.42-6.9 μM) consistently showed two peaks at $s^0_{20,w}$ values of 2.7 S and 4.1 S at all concentrations (Figure 4.2(e,g)). The average $s^0_{20,w}$ values from all speeds and concentrations were 2.52 ± 0.40 S and 4.07 ± 0.14 S (seven

determinations). These values agree with that for monomeric C3d at 3.0 ± 0.1 S and that for unbound CR2 at 3.93 ± 0.20 S, both in 137 mM NaCl (Table 4.1). Integration of the two *c*(*s*) peaks compared to the total peak area revealed 14 ± 6% for the C3d peak and 85 ± 6% for the CR2 peak in the concentration range from 0.06-1.08 mg/ml. This is at the limit of error of the expected ratio of 23%-77% C3d-CR2 calculated according to the masses of the two free proteins. It was concluded that the non-appearance of a *c*(*s*) peak at 4.6 S and the lack of change in signal areas in the two peaks are both consistent with little or weak complex formation in a fast chemical equilibrium on the time scale of the velocity experiment. The K_D value is therefore above 6.9 μM. The use of *c*(*s, f/fo*) distribution fits showed that two frictional ratio values represented the sedimentation profiles at all concentrations (fits not shown). This indicated that any possible weak complex formation did not alter the frictional ratio of CR2 when unbound or bound, i.e. that no conformational change had occurred during complex formation in 137 mM NaCl.

(4.2.5) X-ray scattering of CR2 and the CR2-C3d complex

Solution scattering is a diffraction technique that studies the overall structure of biological macromolecules in random orientations in solution (Perkins *et al*., 2008). Immediately following size-exclusion chromatography, CR2 was successfully studied by X-ray scattering at three concentrations between 0.4 mg/ml and 1.5 mg/ml in 137 mM NaCl. These experiments differ from previous (Gilbert *et al*., 2006b) in that (i) a flow cell was now used, meaning that possible radiation damage effects were more precisely eliminated, and (ii) a new detector had been installed, and enabling lower CR2 concentrations to be studied (Section 4.4.3). However, in contrast to the X-ray experiments performed with 137 mM NaCl, X-ray experiments in 50 mM NaCl were unsuccessful for reason of persistent trace aggregates of CR2 that were visible at the lowest Q values (where $Q = 4\pi \sin\theta / \lambda$; 2θ = scattering angle; λ = wavelength). These aggregates were not visible in the sedimentation experiments because they rapidly sediment out at the start of the velocity runs.

In Guinier plots at low Q values, linear radius of gyration R_G fits in an appropriate $Q.R_G$ range between 0.8 and 1.4 were again obtained (Figure 4.5a). The lowest Q values in Figure 4.5a were not used in order to exclude the effect of unavoidable trace aggregates. The mean X-ray R_G value of CR2 in PBS was 10.8 ± 0.7 nm (11 determinations), which is within error of the previously-reported value of 11.5 ± 0.9 nm (Gilbert *et al*., 2006b). The cross-sectional radius of gyration R_{XS} monitors the

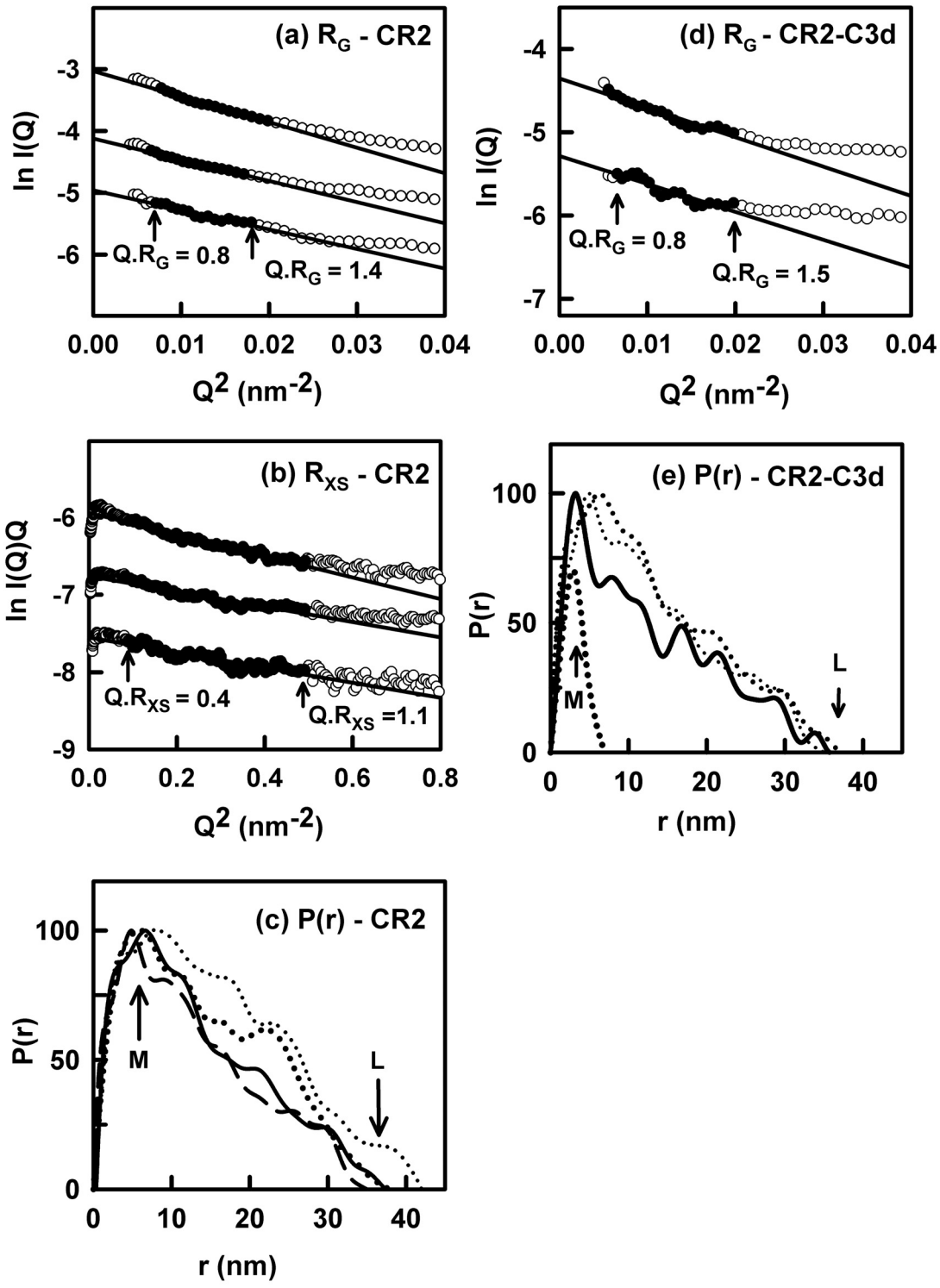


Figure 4.5 X-ray scattering analyses for CR2 and its complex with C3d in 137 mM NaCl.

(a) Guinier R_G analyses, where the filled circles correspond to the data points used to obtain R_G values and the straight lines correspond to the best fit through these points using a Q range of 0.09-0.14 nm^{-1} (top) and 0.08-0.13 nm^{-1} (middle and bottom). The $Q.R_G$ fit range is arrowed. Concentrations of 1.5 mg/ml, 0.8 mg/ml and 0.4 mg/ml

were used (top, middle and bottom respectively).

(b) Guinier R_{XS} analyses, using a Q range of 0.30 - 0.70 nm^{-1}. Other details are as in (a).

(c) Distance distribution $P(r)$ analyses. The intensity of the four $P(r)$ curves are normalised using the maximum of the $P(r)$ curve at M at 6 nm (arrowed). The maximum dimension at 38 nm is denoted by L. The thin dots, line and dashes correspond to 1.5 mg/ml, 0.8 mg/ml and 0.4 mg/ml. The thick dots correspond to 1.84 mg/ml, taken from Gilbert *et al.,* 2006b.

(d) Guinier R_G fit analyses were performed in a Q range of 0.09-0.14nm^{-1}.

(e) Distance distribution $P(r)$ analyses. The maximum in $P(r)$ is denoted by M at 3.2 nm and the maximum dimension is denoted by L at 36 nm. The continuous line corresponds to the CR2 and C3d mixture. The upper pair of dotted lines corresponds to CR2 at 0.8 mg/ml and 0.4 mg/ml from (c). The lowest dotted line corresponds to monomeric C3d in 50 mM NaCl.

shorter axes of an elongated macromolecule. The R_{XS} value monitors the averaged short-range and long-range orientation between adjacent SCR domains. Unlike two other related SCR proteins, factor H and complement receptor type 1, where two linear R_{XS} regions of fit were observed (Aslam & Perkins, 2001; Furtado *et al*., 2008) only one fit range was seen for CR2 (Gilbert *et al*., 2006b). Here, the Q ranges used previously for CR2, factor H and complement receptor type 1 were re-assessed with the new CR2 data at lower concentrations. This reconfirmed the existence of a single R_{XS} region in CR2, and the mean X-ray R_{XS} value of CR2 was determined as 1.5 ± 0.1 nm (nine determinations) (Figure 4.5b), which is similar to the previous value of 1.8 ± 0.1 nm for CR2 (Gilbert *et al.,* 2006b) and other R_{XS} values for SCR proteins (Furtado *et al*., 2008).

The indirect transformation of the scattering curve $I(Q)$ measured in reciprocal space gives the distance distribution function $P(r)$ in real space. The $P(r)$ curve represents all the distance vectors between pairs of atoms within CR2 in PBS. The comparison of the three $P(r)$ curves at 0.4-1.5 mg/ml with that obtained previously at 1.84 mg/ml suggested small concentration dependence in Figure 4.5c that is attributable to trace aggregates at higher concentrations, although the $P(r)$ maximum at M is unchanged at 6 nm and the maximum dimension at L is unchanged at 38 nm.Beyond the small concentration dependence, the finer detail in Figure 4.5c between 5 nm and 35 nm is comparable to the statistical error bars of the $P(r)$ curves, and is not considered significant. However the previously-observed $P(r)$ peak *M2* at 22 nm is not now observed. The mean R_G value of 11 ± 1 nm (3 determinations) calculated from the $P(r)$ curves agrees well with those from the Guinier fits.

The new X-ray data for CR2 was compared with that for the CR2-C3d complex in 137 mM NaCl at three concentrations, for which the sedimentation data in Figure 4.2g shows little or no complex formation had occurred. The mean X-ray R_G value of CR2-C3d was determined to be 11.0 ± 0.2 nm (four determinations) using the same Q range from 0.08 nm^{-1} to 0.14 nm^{-1} (Figure 4.5d). If no complex formation occurs, the weighted mean of the R_G values of 10.8 nm for CR2 and 1.8 nm for monomeric C3d gives 10.4 nm (Gilbert *et al*., 2005; Perkins *et al*., 1985). This value sets the lower limit for the R_G value of a non-interacting mixture of CR2 and C3d. Its similarity with the observed value of 11.0 nm indicates that no or only weak complex formation was detectable. The $P(r)$ curve for CR2-C3d (cotinuous line in Figure 4.5e) corresponds to the sum of those for unbound CR2 at 0.4-0.8 mg/ml and C3 at 0.5 mg/ml (Figure 4.5b in

Gilbert *et al*., 2005). The main peak denoted as M at 3.2 nm is similar to that for C3d, and indicated the presence of unbound C3d. More importantly, the overall length L is unchanged at 38 nm, indicating that any weak interaction between CR2 and C3d does not result in a lengthening of CR2.

(4.2.6) Sedimentation coefficient modelling of CR2, C3d and their complex

The sedimentation coefficients $s^0_{20,w}$ of CR2, C3d and their complex were calculated from their atomic coordinates in order to confirm the interpretation of the sedimentation and X-ray data. The construction of C3d dimers and trimers is necessarily arbitrary as the intermolecular contacts are not known, although two-fold and three-fold symmetry respectively will be present. The compact oligomer arrangements exemplified in Figure 4.6 resulted in calculated $s^0_{20,w}$ values of 2.9-3.1 S, 4.5-4.7 S and 5.8-5.9 S for the monomer, dimer and trimer respectively using the HYDRO and HYDRPRO programs (Section 4.4.2). These calculations agreed well with the experimental values of 3.0 ± 0.1 S, 4.3 ± 0.3 S and 6.0 ± 0.6 S (Table 4.1), and verified the existence of these three forms. The best-fit CR2 model from earlier work (Gilbert *et al*., 2006b) gave calculated $s^0_{20,W}$ values of 3.89-3.94 S, which agreed well with the experimental values of 3.93-4.03 S (Table 4.1). The superimposition of the solution structure of six CR2 SCR 1-2 complexes with C3d (Gilbert *et al*., 2005) onto this best-fit CR2 model using the SCR 2 domain as template gave a calculated $s^0_{20,W}$ value of 4.57 ± 0.04 S for the CR2-C3d complex. This agreed well with the experimental value of 4.52 ± 0.11 S (Table 4.1). The modelling confirmed that no detectable conformational change in unbound CR2 takes place in its complex with C3d.

No X-ray modelling of the CR2-C3d complex was done because this complex did not form in the 137 mM NaCl condition that resulted in good scattering data.

(4.3) Conclusions

We have determined the solution structure for intact CR2 bound to its main physiological ligand C3d to clarify the way in which CR2 interacts with C3d-coated antigens on the B-cell receptors to enhance the antibody response to antigen. To our knowledge, this is the first solution structure for a large multidomain SCR protein bound to its physiological ligand. This study extends our recent CR2 structure which showed that this has an extended and flexible structure, but with evidence of some

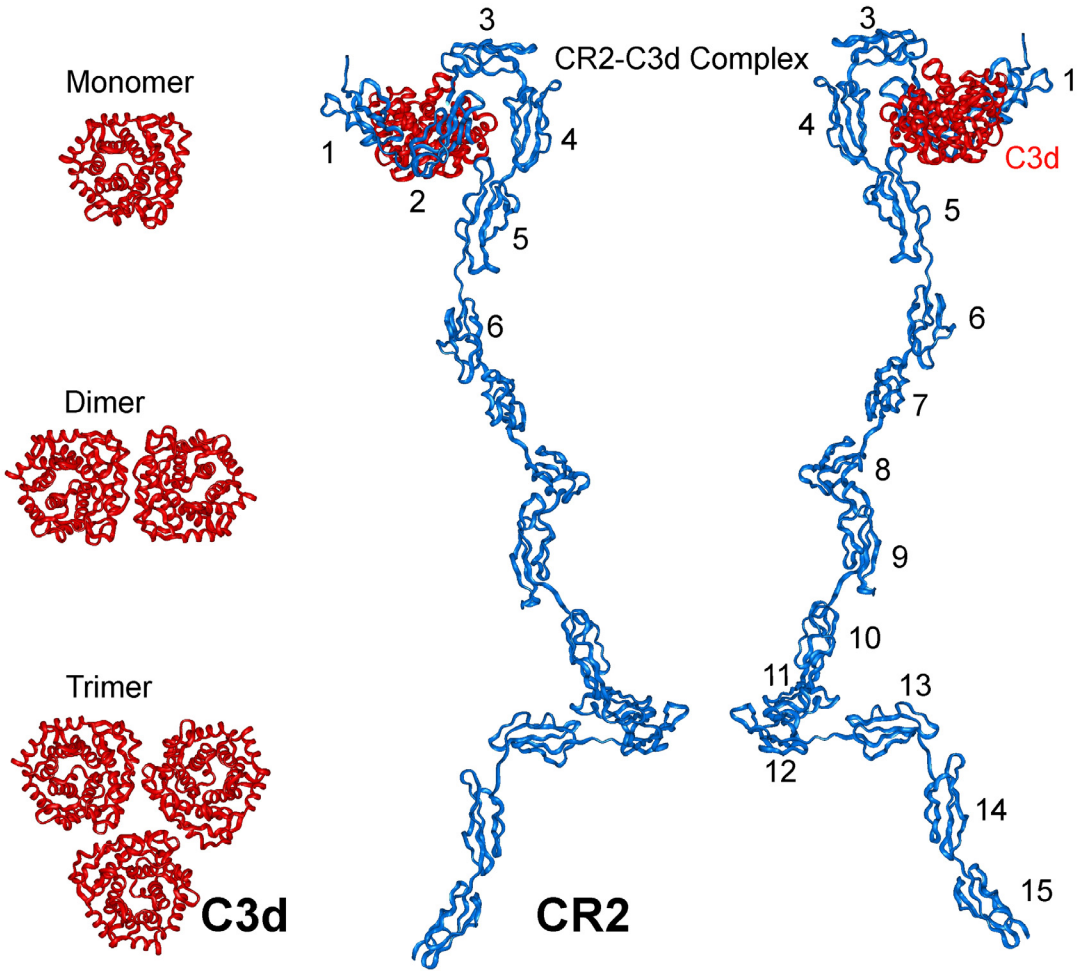


Figure 4.6 Modelling of the CR2-C3d complex using the experimental sedimentation coefficients in 50 mM NaCl. Schematic views of the putative monomer, dimer and trimer forms of C3d are shown in red. The orientation of the monomers within the dimer and trimer is unknown, and is arbitrary as shown. The complex between C3d (red) and CR2 (blue) is shown in both front and back views, in which the 15 SCR domains are numbered from 1 to 15. The sedimentation coefficients calculated from these models are shown in Table 4.1.

bending-back or folding-back of the SCR domains (Gilbert *et al*., 2006b). Our strategy used 50 mM NaCl to clarify molecular details of how the complex was formed, given that no complex was detectable in 137 mM NaCl. Three lines of evidence based on sedimentation coefficients, X-ray scattering, and molecular modelling showed that little conformational alteration, if any, occurs in the 15 SCR domains of CR2 after binding to C3d. This result argues against any allosteric mechanism involving the intact CR2 structure when bound to C3d. Both the unbound and liganded forms of CR2 possess extended and flexible arrangements of SCR domains. This concurs with the proportionately longest set of inter-SCR linker lengths that are present in CR2 when compared with other SCR-containing proteins (discussed more fully elsewhere (Gilbert *et al.,* 2006b; Furtado *et al*., 2008). There are six long linkers of lengths seven or eight residues in CR2. The inter-domain angle between SCR 1-2 is as low as 37^{o}-40^{o} in its crystal structure while solution structures by scattering report angles of 63^{o}-121^{o} (Gilbert *et al*., 2005). This difference is a measure of how much the orientation can vary between two SCR domains in CR2, and results in a flexible SCR arrangement in full-length CR2. Interestingly, our present study indicates no change in the SCR domain arrangement within CR2 with change of ionic strength, showing that this is not influenced by electrostatic interactions between SCR domains.

The present study has extended our understanding of how to study the CR2-C3d interaction. Most notably, C3d in 50 mM NaCl at concentrations above 0.7 mg/ml (20 μM) undergoes two slow monomer-dimer and monomer-trimer associations (possible monomer-tetramer also), both of which disappear in physiological 137 mM NaCl (Figure 4.3). This will hinder any NMR studies of unbound C3d in 50 mM NaCl. If C3d oligomer formation is circumvented for NMR studies by working in 137 mM NaCl, C3d will not form a complex with CR2. Fortunately both types of C3d self-association are blocked when the complex is formed with CR2 in 50 mM NaCl (Gilbert *et al*., 2005). C3d oligomers therefore have no practical consequences on the previous structural analysis of the CR2 SCR 1-2 complex with C3d by constrained X-ray scattering modelling. However, there are further implications for determining the affinity of C3d for CR2 in low salt conditions, as C3d will compete for self-association with itself as well as with CR2. Thus K_D values that are measured in 50 mM NaCl in the absence of an explicit consideration of multiple C3d equilibria should be interpreted with caution. An early binding study of C3d to CR2 in 75 mM NaCl did not fit a simple 1:1 interaction (Sarrias *et al*., 2001). When the interaction between chip-bound CR2 and

fluid phase C3d was studied by surface plasmon resonance, the data suggested one high affinity site and a second much lower affinity site (Asokan *et al*., 2006). Both observations may be explained by the occurrence of C3d dimer and trimer formation. In addition, it cannot be ruled out that earlier C3d-CR2 mutagenesis work that modified C3d may have affected C3d dimer and trimer formation, rather than the direct C3d interaction with CR2. In distinction to these single-molecule studies, comparison with cell systems shows a detectable interaction between CR2 and C3d at physiological ionic strength. Thus flow cytometry using tetrameric C3dg-biotin in solution bound to CR2-transfected cells bind the C3dg ligand in PBS buffer in a dose-dependent manner (Hannan *et al*., 2005). This difference is attributable to the involvement of a large number of CR2 molecules at the cell surface, and shows that solution results based on single molecule interactions do not necessarily correspond to *in vivo* conditions.

The loss of the CR2-C3d interaction in 137 mM NaCl (Figure 4.2) may be explained by the masking of electrostatic interactions that mediate complex formation in 50 mM NaCl. As noted above, possible salt-dependent conformational changes within the SCR domains that might inhibit the CR2 interaction with C3d have been ruled out. This key observation must be added to determinations of the dissociation constant K_D for the CR2-C3d complex in five earlier studies by surface plasmon resonance (Table 4.2). It should be noted that the ultracentrifugation data provide an independent method for determining associations, as this corresponds to a fluid phase measurement, not one based on binding to a chip surface. In low salt conditions of 75 mM NaCl and less, only two of the five studies report K_D values below 100 nM (Guthridge *et al.,* 2001; Hannan *et al*., 2005). The ultracentrifugation data in Figure 4.2h clearly show that a K_D of less than 280 nM will account for the *c(s)* plot, and these support the lowest K_D values in Table 4.2. Given that the largest K_D values in low NaCl in Table 4.2 mostly involve CR2 SCR 1-15, it is possible that these larger K_D values may be explained by a lack of steric access to the SCR 1-2 binding site in CR2 caused by the high flexibility of the SCR 3-15 domains. In contrast, using high salt conditions, all three surface plasmon resonance studies and the present ultracentrifugation study show that the K_D values are significantly higher at 600 nM and above (Table 4.2). All four studies confirm the same weakening of C3d binding when the [NaCl] is increased.

The molecular basis of NaCl effects on the CR2-C3d association is informed using crystal and solution structures. The CR2 SCR 1-2 structure was obtained in 100 mM $MgCl_2$,

Table 4.2 Dissociation constants K_D for the CR2-C3d complex.

CR2 protein	Ligands	Buffer	Experimental K_D	Reference
Low NaCl (~50 mM)				
CR2 SCR 1-15	C3dg	10 mM phosphate, 50 mM NaCl	400-1600 nM	(Moore *et al*., 1989)
CR2 SCR 1-15 BCCP	C3d	20 mM Phosphate, 50 mM NaCl	4300 nM	(Sarrias *et al*., 2001)
CR2 SCR 1-15 BCCP	C3d	20 mM Phosphate, 75 mM NaCl	2800 nM	(Sarrias *et al*., 2001)
CR2 SCR 1-2	C3dg-biotin	10 mM HEPES, 50 mM NaCl	22.4 nM*	(Guthridge *et al*., 2001)
CR2 SCR 1-15	C3dg-biotin	10 mM HEPES, 50 mM NaCl	27.1 nM	(Guthridge *et al*., 2001)
CR2 SCR 1-2	C3d	10 mM HEPES, 50 mM NaCl	67.2 nM	(Hannan *et al*., 2005)
CR2 SCR 1-4 biotin	C3d	10 mM HEPES, 50 mM NaCl	314 nM	(Asokan *et al*., 2006)
CR2 SCR 1-2	C3d-biotin	10 mM HEPES, 50 mM NaCl	179 nM	(Asokan *et al*., 2006)
CR2 SCR 1-4	C3d-biotin	10 mM HEPES, 50 mM NaCl	617 nM	(Asokan *et al*., 2006)
CR2 SCR 1-15	C3d-biotin	10 mM HEPES, 50 mM NaCl	2200 nM	(Asokan *et al*., 2006)
CR2 SCR 1-15	C3d	10 mM HEPES, 50 mM NaCl	Below 280 nM	Present study (Figure 4.2h)
High NaCl (> 100 mM)				
CR2 SCR 1-15	C3dg	10 mM phosphate, 100 mM NaCl	3200 nM	(Moore *et al*., 1989)
CR2 SCR 1-2	C3dg-biotin	10 mM HEPES, 125 mM NaCl	620 nM**	(Guthridge *et al*., 2001)
CR2 SCR 1-2	C3dg	10 mM HEPES, 125 mM NaCl	658 nM	(Hannan *et al*., 2005)
CR2 SCR 1-15	C3d	9.6 mM phosphate, 137 mM NaCl	Above 6900 nM	Present study (Figure 4.2g)

* No change in K_D was observed between 25 mM NaCl to 85 mM NaCl.

** Precise values in 125 mM NaCl could not be obtained for reason of signal-noise issues at high [NaCl].

100 mM CsCl, 100 mM HEPES (Prota *et al*., 2002). The co-crystal structure was obtained using 200 mM zinc acetate, 100 mM sodium cacodylate (Szakonyi *et al*., 2001). These buffers are probably sufficient to dissociate SCR1 from C3d according to our data (Figure 4.2). A recent study of the effect of zinc ions on the binding of CR2-IgG to plate-bound C3dg showed that the interaction is abrogated in the presence of 0.2 M zinc acetate (Isenman *et al*., 2010). In the co-crystal, only SCR2 associates with C3d through the hydrogen bonding of the R83 side chain with a pocket of main-chain carbonyl groups, as well as through two lysines. The importance of R83 on C3d for binding with CR2 was indicated in a mutation study (Hannan *et al*., 2005) and a chemical shift perturbation study (Kovacs *et al*., 2009); however, the hydrogen bonds were disputed as an artefact in view of an alanine mutagenic scan (Clemenza & Isenman, 2000, see also Isenman *et al*., 2010). In the solution state, constrained scattering modelling shows that both SCR structures in uncomplexed CR2 SCR1-2 are opened out in 10 mM HEPES, 50 mM NaCl. This approach, as well as mutagenesis studies in 50 mM NaCl and 137 mM NaCl, also showed that both SCR 1 and SCR 2 make contact with the surface of C3d in this complex (Gilbert *et al*., 2005; Hannan *et al*., 2005; Young *et al*., 2007; Isenman *et al*., 2009). Thus a single SCR2-C3d interaction on its own in 137 mM NaCl appears to be too weak to lead to complex formation (Figure 4.2), and both SCR domains appear to be needed for binding.

Electrostatic calculations have been reported for the CR2-C3d complex (Morikos & Lambris, 2004; Zhang *et al*., 2007), but their utility appears limited by the consideration of only the C3d-SCR2 interaction in the complex. Interestingly the calculations do not explain the absence of CR2-C3d complex formation in 137 mM NaCl, e.g. through the prediction of p*K* changes that would lead to the loss of charges that would lead to the loss of charges at the CR2-C3d interface. The calculations show that C3d has an excess charge of -4, while CR2 SCR 1-2 has an excess charge of +8. The calculations postulate that long-range electrostatic interactions have a role in complex formation, and it is possible that their masking in physiological salt is responsible for complex dissociation. Electrostatic maps for CR2 SCR 1-2 show a preponderance of positive charge on its surface (Hannan *et al*., 2005; Young *et al*., 2007). In C3d two separate clusters of residues on opposite sides of the acidic pocket, specifically D37/D39 and E160/D163/I164/E166 are found to be important for binding with CR2 (Clemenza & Isenman, 2000; Isenman *et al*., 2010). Unlike the crystal complex of CR2-C3d, CR2 mutations targeting R13, R28, R36, K41, K50 and K57 on SCR1 confirm that SCR1 interacts with C3d, probably with at least some of E117,

D122, D128 and D147, which form a negatively-charged groove on C3d adjacent to the SCR2 site (Hannan *et al*., 2005). This interaction is supported by the scattering modelling of this complex (Gilbert *et al*., 2005). These short range interactions stabilise the complex, but are not sufficient on their own to be effective in 137 mM NaCl to enable the complex to be formed on the basis of a 1: 1 single molecule interaction. It may be relevant that the mutagenesis work used a tetrameric form of C3dg-biotin ligated to streptavidin, suggesting that the aggregation of CR2 and C3d will promote these interactions in physiological salt.

The understanding of the solution structures of CR2 and the CR2-C3d complex is related to its involvement with the B-cell receptor as a coreceptor that links innate and adaptive immunity (Fearon & Carroll, 2000; Tolnay & Tsokos, 1998). Our previous study (Gilbert *et al*., 2006b) was focussed on the determination of a highly flexible CR2 structure that is well designed to interact with a variety of differently-sized locations for C3d bound to antigens. This CR2 structure is conserved across different mammalian species. The present study has focussed on the solution-phase structure of the CR2-C3d complex and its reduced stability in physiological salt. Here and earlier (Moore *et al*., 1989; Guthridge *et al*., 2001) no interaction was detected between C3d and CR2 in physiological pH and ionic strength. Nonetheless, the use of a tetravalent C3dg ligand bound to surface-bound CR2 shows that this interaction must take place. Our main results show that B-cell activation must result from the combination of (i) a flexible CR2 domain structure that is able to position SCR 1-2 in conformations appropriate to interact with antigens, and (ii) the accumulation of weak interactions between CR2 and C3d. For the latter to contribute as an immunologically significant trigger, these weak CR2–C3d interactions must be multiplied several times over by the presence of a sufficiently large number of C3d-bound ligands on the antigen(s)-coated pathogen and a sufficiently large number of CR2 molecules on the B-cell surface. This interaction will be amplified by increased local concentration effects at the B-cell surface, because CR2 is already positioned on this surface ready for interaction with C3d ligands. Such an interaction in physiological salt appears to be a good way for B-cells to respond only to antigens that present multimeric C3d molecules in a manner that clustered surface-bound CR2 molecules. This is similar to the CR1 interaction with C3b/C4b molecules on neutrophils, which is also strongly promoted by the polymerization or multimerization of the ligand and the clustering of receptors.

(4.4) Materials and Methods.

(4.4.1) Purification of CR2, C3d and their complex

Human CR2 SCR 1-15 was expressed in Sf9 insect cells as a soluble recombinant protein in the baculovirus expression system as described previously (Guthridge *et al*., 2001; Gilbert *et al*., 2006b). This protein was kindly provided by Dr Asokan (Denver, U.S.A). This product (Ile21 to Arg971) possessed a stop codon prior to the transmembrane domain to create a soluble secreted form of the receptor. Immediately prior to scattering and ultracentrifugation data acquisition, the concentrate was dialysed against Dulbecco's phosphate-buffered saline (PBS) (137 mM NaCl, 2.7 mM KCl, 8.1 mM Na_2HPO_4, 1.5 mM KH_2PO_4, pH 7.5) made 0.5 mM in EDTA. Non-specific aggregates were removed by size exclusion chromatography on a Superdex 200 HiLoad 16/60 column (Amersham Pharmacia Biotech).

Human C3d (sequence 994-AVDAE… to …QLPSR-1303 of the mature protein C3 with SWISSPROT accession code P01024) was expressed in *E. coli* BL21 transformed with the PGEX-2T vector for purification by an affinity column binding GST-tag using standard protocols. The construct was provided by Prof. V. M. Holers and Dr. J. P. Hannan of the University of Colorado School Of Medicine, Denver, Colorado, U.S.A. Previously C3d was purified using a pET15b-C3d plasmid (Szakonyi *et al*., 2001; Guthridge *et al*., 2001; Gilbert *et al*., 2005). Ampicillin-resistant colonies were used to produce starter cultures, which were then expanded to five litres and grown at 37°C until an absorbance at 600 nm of 0.3 was attained, then cooled to 23°C and incubated to an absorbance of 0.4. Samples were then induced with 0.5 mM IPTG and shaken overnight at 25°C. The culture was lysed in the presence of protease inhibitors by sonication, and after clarification applied to a GSTrap FF column equilibrated in 25 mM Tris, 0.5 M NaCl, 1 mM EDTA, 1 mM DTT, pH 8.0. After washing to remove unbound protein, thrombin was added to the column. After an overnight incubation at 4ºC, the thrombin was inactivated and the cleaved C3d was eluted by further washing. Reduced glutathione was added to the buffer in further washes to elute GST and uncleaved GST-C3d. The C3d was concentrated using Amicon® Ultra-15 centrifugal filter devices with a molecular mass cut-off of 10 kDa at 4500 *g*, and size exclusion chromatography using a Superdex-200 Prep Grade XK 16/60 column resulted in the C3d samples used for data acquisition.

The CR2 and C3d samples were extensively dialysed into PBS or HEPES (10 mM HEPES, 50 mM NaCl, pH 7.4) buffers for data acquisition. Samples were routinely analysed by SDS-PAGE before and after scattering and ultracentrifugation to confirm their integrity. CR2 and C3d concentrations were determined using absorption coefficients of 12.7 and 13.1, respectively (1%, 280 nm, 1 cm path length), which was calculated from their compositions (Perkins, 1986). For the CR2 composition, eleven high mannose type oligosaccharides $GlcNAc_2Man_7$ were assumed. Following dialysis into PBS or HEPES buffer as required, the CR2-C3d complexes were prepared by mixture in a 1:1 molar ratio according to the sample absorbances.

(4.4.2) Analytical ultracentrifugation data and analyses

Analytical ultracentrifugation was done at 20ºC with two Beckman XL-I instruments equipped with An50Ti and An60Ti rotors as described (Gilbert *et al.*, 2006b). Sedimentation velocity experiments were used for shape and size analysis. New sedimentation velocity data for CR2 were acquired in both PBS and Hepes buffers at rotor speeds of 25,000, 30,000, 35,000, 40,000 and 42,000 r.p.m in a concentration series of 0.12 – 0.34 mg/ml. New sedimentation velocity data for C3d were acquired in both PBS and Hepes buffers. The PBS data utilised sample concentrations of 0.54 – 7.48 mg/ml, while the Hepes data utilised concentrations of 0.32 – 8.29 mg/ml, with rotor speeds of 32,000, 40,000 and 50,000 r.p.m. This work included reanalysis of the previous C3d velocity data in Hepes buffer (Gilbert *et al.*, 2005). Sedimentation velocity experiments for the CR2–C3d complex were performed with rotor speeds of 30,000, 40,000 and 42,000 r.p.m. in PBS and Hepes buffers. The PBS data used a total concentration of complex of 1.08 mg/ml, prepared by mixing 1.07 mg/ml CR2 and 1.13 mg/ml C3d in a 1:1 molar ratio. The Hepes data used a total concentration of 0.54 mg/ml, prepared by mixing 0.67 mg/ml CR2 and 0.33 mg/ml C3d in a 1:1 molar ratio. Each complex was studied in a dilution series from 100% to 10%. Data analyses used direct boundary Lamm fits of up to 200 scans using SEDFIT (Schuck, 1998, 2000), which was used for size-distribution analysis $c(s)$ that assumed that all species have the same frictional ratio f/f_o. The final $c(s)$ fits were determined using a resolution of 200 and by floating the meniscus, frictional ratio, and baseline and holding the cell bottom, partial specific volume and solvent density fixed until the overall root-mean-square deviations and visual appearance of the fits were satisfactory (Figure 4.2 and 4.3). For C3d, the frictional ratio was floated to result in a realistic range of values of 1.10 ± 0.08 (Table 4.1). For CR2 and the CR2–C3d complex, floating the frictional ratio resulted in

ranges of 1.99 – 2.11 (Table 4.1). The buffer density of 137 mM NaCl in PBS and 50 mM NaCl in Hepes was 1.00541 g/ml and 1.00117 g/ml, respectively, as measured with an Anton Paar DMA 5000 density meter, to be compared with 1.00534 ml/g and 1.00030 ml/g, respectively, calculated from SEDNTERP (Gilbert *et al*., 2005). The partial specific volume *v* for CR2 and C3d was calculated to be 0.716 ml/g and 0.746 ml/g from its amino acid and carbohydrate content calculated from its sequence, respectively (Perkins, 1986).

Sedimentation equilibrium experiments resulted in molecular weight and dissociation constant K_D. C3d data were obtained in PBS and HEPES buffers at 20ºC over 30 h at each speed using six-sector cells with column heights of 2 mm and rotor speeds of 19,000 r.p.m., 24,000 r.p.m., 28,000 r.p.m., 30,000 r.p.m. and 35,000 r.p.m. The existence of equilibrium conditions after at least 10 h in PBS buffer and at the lowest concentrations, after at least 30 h at the highest concentrations in HEPES buffer was confirmed using overlays of runs measured at 3 h intervals in SEDFIT. In PBS buffer, nine concentrations between 0.04 mg/ml to 7.01 mg/ml were used, while in HEPES buffer, nine concentrations between 0.11 mg/ml to 5.50 mg/ml were used. Data analyses utilised SEDPHAT version 5.01 (Schuck, 2003). First, multispeed fits assuming a single species were performed individually. Next, the best interference optics fits from 19,000 r.p.m., 28,000 r.p.m., and 35,000 r.p.m. were used for global fits using the monomer-'m-mer'-'n-mer self-association model with m = 2 (dimer) and n = 3 (trimer) in SEDPHAT (Figure 4.4). The global fits utilised a common baseline, meniscus, cell bottom, extinction coefficient, and loading concentration. Fits used the M and s, simplex and Marquardt-Levenberg fitting routines, while floating the baseline and the cell bottom of the cell. Statistical error analyses employed the Monte-Carlo approach.

(4.4.3) X-ray scattering data and analyses

X-ray scattering data were collected in three sessions on the camera at the ID02 high brilliance beamline at the European Synchrotron Radiation Facility in Grenoble, France, operating at a ring energy of 6.03 GeV (Narayanan *et al*., 2001). Data prior to September 2006 (Gilbert *et al*., 2005, 2006a, 2006b) were acquired using an image-intensified FReLoN CCD detector, which involved an additional correction step for the point spread function of the image intensifier. This detector has been replaced by a fibre optically coupled high sensitivity and dynamic range CCD detector (FReLoN). Using a smaller beamstop, this enabled data to be recorded at a shorter sample-to-detector

distance of 2.0 m, thereby improving the signal-to-noise ratio. For CR2 in PBS buffer, storage ring currents ranged between 167-168 mA (uniform fill mode) and 79-81 mA (16 bunch mode) in two sessions. For the CR2-C3d complex in PBS buffer, the ring current ranged from 56 mA to 82 mA (16 bunch mode in one beam session). Flow cells were used to eliminate radiation damage in conjunction with on-line checks for the absence of radiation damage by optimisation of the exposure times in sets of ten time frames. The flow cell utilised fresh 100 µl samples that move in the X-ray beam, in distinction to the static samples in Perspex cells used previously (Gilbert *et al*., 2005, 2006a, 2006b). Each sample was measured in four sets of ten time frames of length 0.1, 0.15 and 0.2 sec for CR2 and 1, 1.5 and 2 sec for the CR2-C3d complex, both in PBS. Sample temperature corresponded to ambient conditions at 23ºC for CR2 and 20ºC for the CR2-C3d complex. Unbound CR2 was studied at concentrations of 0.4 mg/ml, 0.8 mg/ml and 1.5 mg/ml. The CR2-C3d complex was studied at 0.7 mg/ml, 1.0 mg/ml and 1.3 mg/ml. Buffers were measured using the same exposure times in alternation with the samples to eliminate background subtraction errors.

For a given solute-solvent contrast, the radius of gyration R_G characterises the degree of elongation provided that the internal inhomogeneity of scattering densities has no effect. Guinier analyses at sufficiently low Q give the R_G and $I(0)$ parameters (Glatter & Kratky, 1982):

$$\ln I(Q) = \ln I(0) - R_G^2 Q^2/3$$

This equation is valid in a $Q.R_G$ range of up to 1.5. The relative $I(0)/c$ values (where c is sample concentration) for samples measured in the same buffer gives the relative molecular mass of the proteins when referenced against a suitable standard. For elongated structures, the mean radius of gyration of cross-sectional structure R_{XS} and the mean cross-sectional intensity at zero angle $[I(Q)Q]_{Q\to 0}$ is obtained from:

$$\ln [I(Q)Q] = [I(Q)Q]_{Q\to 0} - R_{XS}^2 Q^2/2.$$

Combination of the R_{XS} and R_G analyses gives the length of the macromolecule if the structure can be represented by an elliptical cylinder, where $L = \sqrt{12}(R_G^2 - R_{XS}^2)$ (Glatter & Kratky, 1982). The Guinier analyses were done with a PERL script program SCTPL7 (J. T. Eaton and S. J. Perkins, unpublished software) on Silicon Graphics O2 Workstations. Indirect transformation of the reciprocal space scattering data into real space to give the distance distribution function $P(r)$ was implemented using the program GNOM (Semenyuk & Svergun, 1991; Svergun, 1992). The $P(r)$ curve represents the

distribution of distance vectors r between pairs of atoms within a molecule, and provides another calculation of the R_G and $I(0)$ values as well as giving the maximum dimension of the macromolecule, L. The CR2 X-ray curve used the Q range between 0.08 nm^{-1} and 1.75 nm^{-1} at 0.40 mg/ml (Figure 4.5c), supplemented by that between 0.09 nm^{-1} and 1.34 nm^{-1} at 0.80 mg/ml and between 0.09 nm^{-1} and 1.69 nm^{-1} at 1.5 mg/ml. The CR2-C3d complex X-ray curves use the Q range between 0.12 nm^{-1} and 1.45 nm^{-1} at 1.32 mg/ml (Figure 4.5e), and Dmax was set as 36 nm.

(4.4.4) Modelling of the CR2-C3d complex

The CR2 SCR 1-15 model was taken from the overall best-fit solution structure of CR2 which gave good agreement for both the R_G and R_{XS} values (PDB code 2gsx) (Gilbert *et al*., 2006b). That for the complex between CR2 SCR1-2 and C3d was taken from the six best-fit solution structures for this (PDB code 1w2s) (Gilbert *et al*., 2005). Insight II 98.0 (Accelrys, San Diego, CA) software on Silicon Graphics workstations was used for manipulations. Thus the C3d oligomers were created by juxtapositions of the C3d crystal structure (PDB code 1C3d) (Nagar *et al*., 1998) in arbitrary contacts but showing two-fold or three-fold symmetry. SCR2 in the six CR2-C3d complexes was superimposed upon SCR2 in the SCR1-15 model, after which the SCR1 and SCR2 domains in the SCR 1-15 model were deleted. Minor steric overlap between the C3d fragment and SCR3-15 in these models was removed by a torsion angle rotation in the linker peptide between SCR2 and SCR3. Sphere models for C3d were calculated as described for use with the HYDRO software (Gilbert *et al*., 2005; Garcia de la Torre *et al*., 1994). Sedimentation coefficients for CR2, C3d and their complex were calculated from atomic models using HYDROPRO software with a default value of 0.31 nm to represent the hydration shell (Garcia de la Torre *et al*., 2000).

Chapter Five

Self-association and domain rearrangements between complement C3 and C3u provide insight into the activation mechanism of C3

(5.1) Introduction

The complement system of innate immunity plays a major role in the recognition and elimination of microbial intruders and other pathogenic cells (Walport, 2001; Law & Reid, 1995). Complement is activated by three pathways, the classical, alternative and lectin pathways, each of which leads to the activation of C3, the so-named third component of complement. C3 is the most abundant complement component in plasma at about 1.0 mg/ml, and its level can be elevated considerably during inflammation and infection. The three pathways lead to a C3 convertase enzyme complex that activates C3 by proteolytic cleavage to yield the small anaphylatoxin C3a and active C3b (Walport, 2001; Law & Reid, 1995). The removal of C3a induces a major conformational change in C3b to expose an internal thioester bond, which is inaccessible in C3 (Janssen *et al*., 2005; Fredslund *et al*., 2006, 2008; Janssen *et al*., 2006; Wiesmann *et al*., 2006; Wu *et al*., 2009; Rooijakkers *et al*., 2009; Ajees *et al*., 2006). The thioester forms covalent bridges with antigenic surfaces. C3 is also continuously activated at a slow "tick-over" rate in plasma to form C3u, sometimes named $C3_{H2O}$. C3u also participates in convertase enzyme complexes that cleave C3 into C3a and C3b, and triggers the rapid activation of C3 to C3b, but C3u is unable to attach covalently to surfaces because its thioester bond has been hydrolysed.

C3 is synthesised as a single chain pre-pro-molecule in which the α- and β-chains are linked by a tetra-Arg sequence that is enzymatically removed during post-translational processing. The α- and β-chains (molecular masses 115 kDa and 75 kDa respectively) are linked by a single disulphide bond and non-covalent interactions. Major aspects of complement activation have been revealed by three C3 and C5 crystal structures (where C5 is a homologue of C3) and five C3b crystal structures (Janssen *et al*., 2005; Fredslund *et al*., 2006, 2008; Janssen *et al*., 2006; Wiesmann *et al*., 2006; Wu *et al*., 2009; Rooijakkers *et al*., 2009; Ajees *et al*., 2006). C3 is formed from a compacted arrangement of eight macroglobulin (MG) domains, to which are added the thioester-containing domain (TED) domain containing the C3 active site, a C1r/C1s-Uegf-Bmp1 (CUB) domain that links the TED and MG domains, and three additional C345C, L and C3a domains. In C3, the TED and CUB domains are positioned near the C345C domain and the thioester is buried (Janssen *et al*., 2005; Fredslund *et al*., 2006, 2008). In C3b, four similar crystal structures show that the TED and CUB domains are extended along the major axis of C3b to make contact with the MG1 domain at the base of the structure (Janssen *et al*., 2006; Wiesmann *et al*., 2006; Wu *et al*., 2009;

Rooijakkers *et al*., 2009). A fifth deviant crystal structure suggested that the TED and CUB domains are significantly extended away from the MG domains (Ajees *et al*., 2006), however the crystallographic evidence for the fifth C3b structure was disputed (Janssen *et al*., 2007; Ajees *et al*., 2007; Borrell, 2009). Electron microscopy studies of C3b *in vacuo* supported the location of the TED and CUB domains seen in the four C3b crystal structures (Nishida *et al*., 2006). H-D exchange rates that were measured in mass spectrometry studies of C3, C3u and C3b indicated that large conformational changes occur between the three different protein forms (Winters *et al*., 2005; Schuster *et al*., 2008).

Knowledge of the solution properties of C3 and C3u will clarify the initial stage of the mechanism of the conversion of C3 to C3b. A notable aspect of these crystal structures is that many were crystallised in buffers with low salt, and it is not known if these crystallographically-observed C3 structures or their properties are maintained in near-physiological buffers. A two-segmented structure for C3 and its homologues C4 and C5 had been identified from low resolution solution scattering modelling, in which large conformational differences between C4 and C4b were detectable (Perkins & Sim, 1986; Perkins *et al*., 1990a; 1990b). Because the smaller segment seen by scattering resembles the distinct TED-CUB domains seen in recent crystal structures, this indicates that scattering modelling is able to monitor the positions of the TED-CUB domains in C3 and C3u (Figure 5.1). In recent years, X-ray scattering has benefitted from improved signal-noise ratios at the ESRF facility, alongside with new constrained modelling methods that result in molecular structural determinations that are deposited in the Protein Data Bank (Bonner *et al*., 2009; Perkins *et al*., 2009). Analytical ultracentrifugation has been much improved by the use of size-distribution $c(s)$ analyses to analyse self-association (Cole *et al*., 2008). Here, we combine both methods to show that both self-association and conformational differences exist between C3 and C3u. We show that C3 and C3u are principally monomeric, however the monomer exists in equilibria with two different types of dimers. One of these offers an explanation for the mechanism of inhibition of C3b by the bacterial protein SCIN (Rooijakkers *et al*., 2009). Constrained scattering modelling demonstrated that the position of the TED-CUB domains relative to the MG1-MG8 domains is significantly altered between C3 and C3u. We conclude that the conformational properties of the TED-CUB domains are more variable than believed from crystallography. This variability would enable activated C3

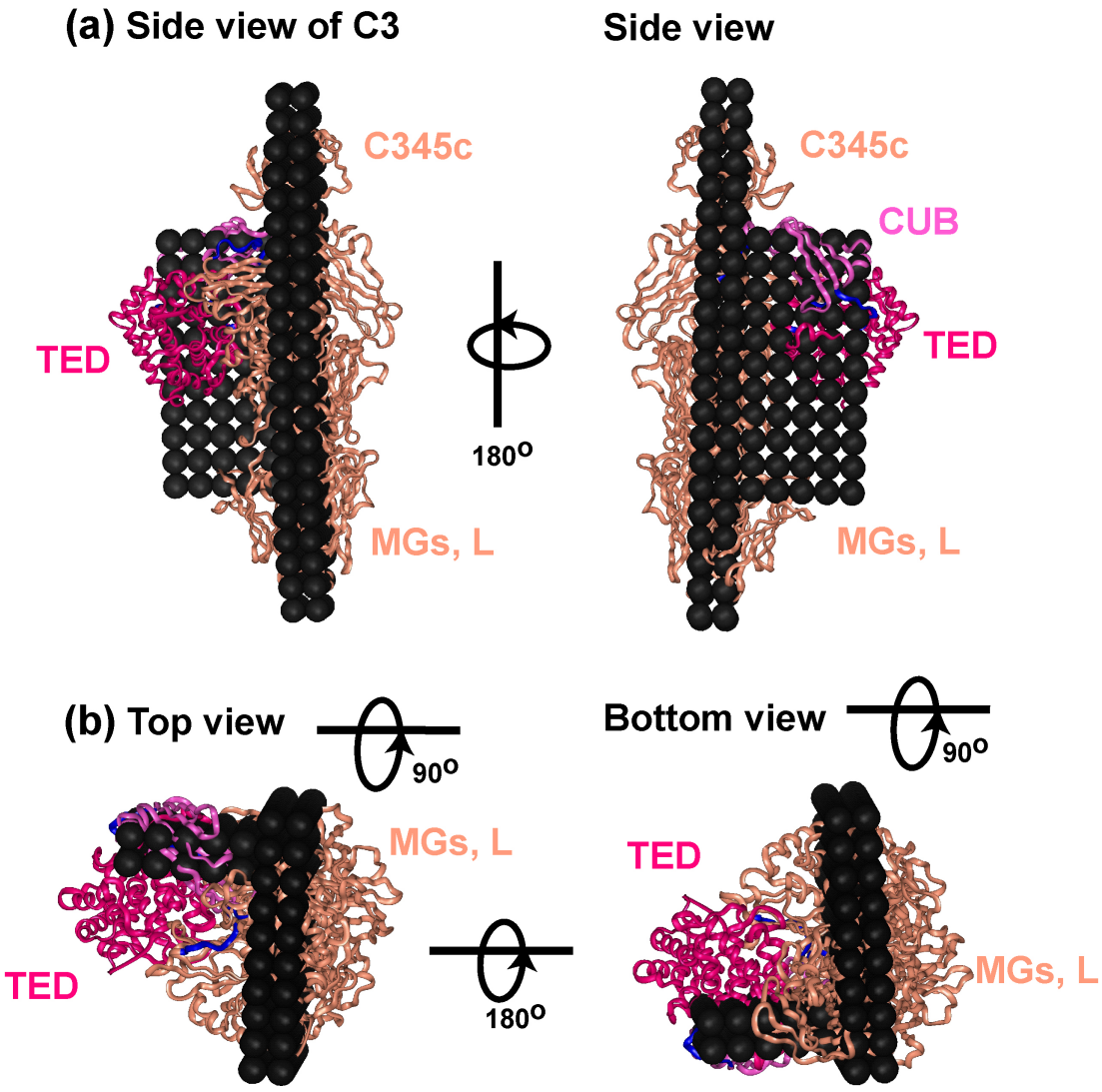


Figure 5.1 Comparison between the low resolution Debye sphere models for C3, C4 and C5 in 1990 and the crystal structure of C3 (PDB code 2A73).

(a) The two side views correspond to orthogonal views of the long axis of the C3 structure. The long axis of the C3c segment (tangerine) in the C3 crystal structure corresponds to the major segment of the C3 sphere model.

(b) The top end of the two views in (a) is rotated by 90° towards the reader. The TED-CUB domains (red) correspond to the minor segment of the C3 Debye sphere model.

and its homologues to bind to a broad range of antigenic targets and provides insight for the mechanism of C3 regulation by complement factor H and other inhibitors.

(5.2) Results

(5.2.1) Sedimentation velocity data analysis for C3 and C3u

Immediately prior to ultracentrifugation and scattering experiments, purified C3 and C3u (Section 5.4.1) were subjected to gel filtration on Superose 6 column to remove potential pre-existing aggregates or degradation products. Both proteins eluted as single symmetric peaks in 137 mM NaCl and 50 mM NaCl (Figures 5.2(a,b)). Here and below, the buffer is denoted as 137 mM NaCl or 50 mM NaCl, even though phosphate is also present (Section 5.4). C3u (and less so with C3) eluted earlier in 50 mM NaCl than in 137 mM NaCl, suggesting that C3u underwent self-association in reduced salt. C3 and C3u before and after experiments revealed a single band in non-reduced SDS-PAGE and two bands corresponding to the α- and β-chains in reduced SDS-PAGE (Figure 5.2(c)).

Analytical ultracentrifugation studies macromolecular structures in solution by monitoring the time-course of sedimentation under high centrifugal force (Cole *et al.*, 2008). The peaks in size distribution analyses $c(s)$ plots monitor the extent of sample monodispersity, and macromolecular elongation is monitored through the sedimentation coefficient $s_{20,w}$ values of the observed peaks. Using SEDFIT, good boundary fits were obtained for the sedimentation profiles for each of C3 and C3u in 50 mM and 137 mM NaCl buffer (Figure 5.3(a-d)). While a major $c(s)$ peak was visible in all runs, differences between the two buffers were seen:

(i) In 137 mM NaCl, the interference and absorbance data sets for C3 gave an $s^0_{20,w}$ value of 7.85 ± 0.05 S, and both data sets for C3u gave an $s^0_{20,w}$ value of 7.44 ± 0.07 S (Figure 5.4). No concentration dependence was seen, showing that both the shape and size of C3 and C3u were each unchanged with concentration, and the molecular weights from the $c(s)$ peaks corresponded to 192 ± 8 kDa and 172 ± 16 kDa for C3 and C3u respectively, in good agreement with their sequence-predicted molecular weight of 189.0 kDa. Accordingly the change in $s_{20,w}$ values showed that C3u is more elongated in solution than C3. The frictional ratios f/f_o (where f_O is the frictional coefficient of the sphere with the same volume as the hydrated glycoprotein) for C3 and C3u were 1.37 ± 0.10 and 1.40 ± 0.10.

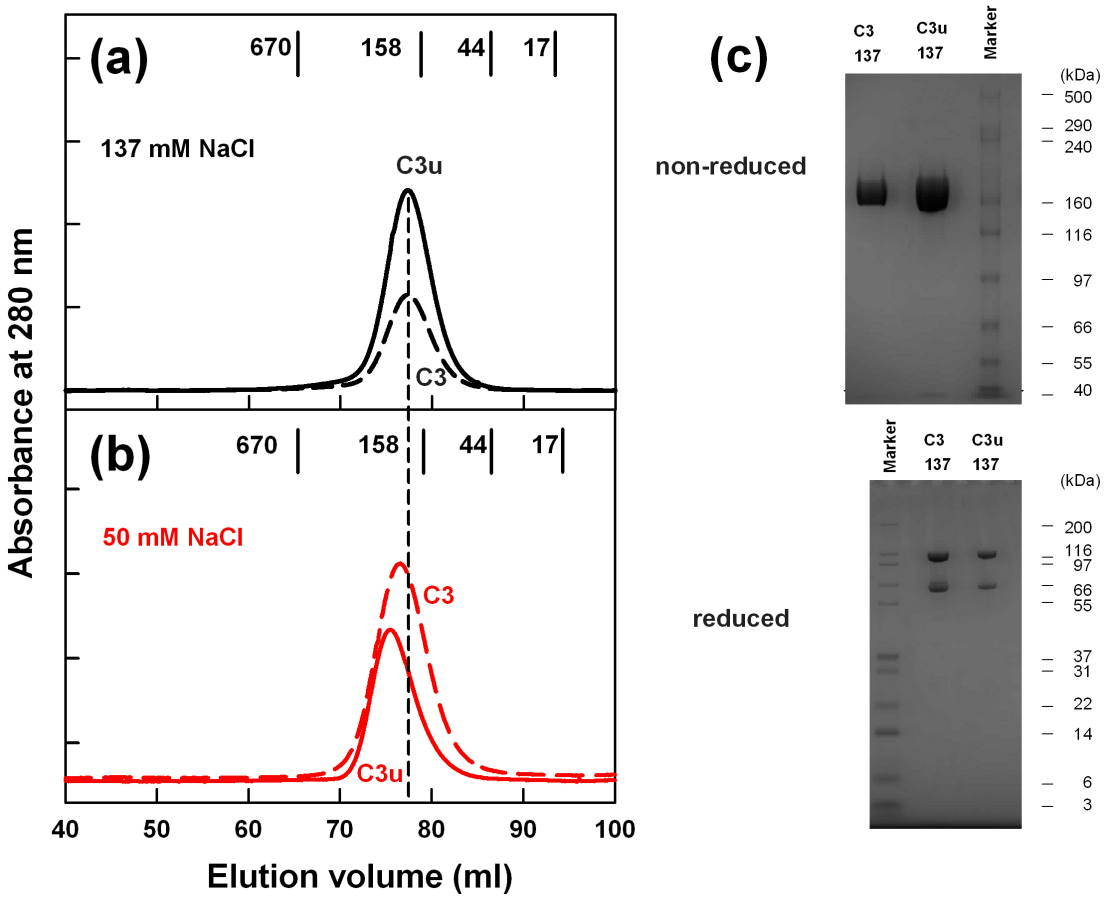


Figure 5.2 Size-exclusion gel filtration and SDS-PAGE analyses of C3 and C3u.

(a,b) The elution of C3 (dashed line) and C3u (continuous line) are labelled as shown. The elution positions of four standards are shown in kDa. In 137 mM NaCl buffer (black), C3 and C3u eluted as a single peak at approximately the same position. In 50 mM NaCl buffer (red), C3u eluted slightly earlier than C3, and both eluted earlier compared to 137 mM NaCl buffer.

(c) Non-reducing and reducing SDS-PAGE analyses of C3 and C3u from the 137 mM NaCl buffer. The protein markers are labelled at the right side in kDa.

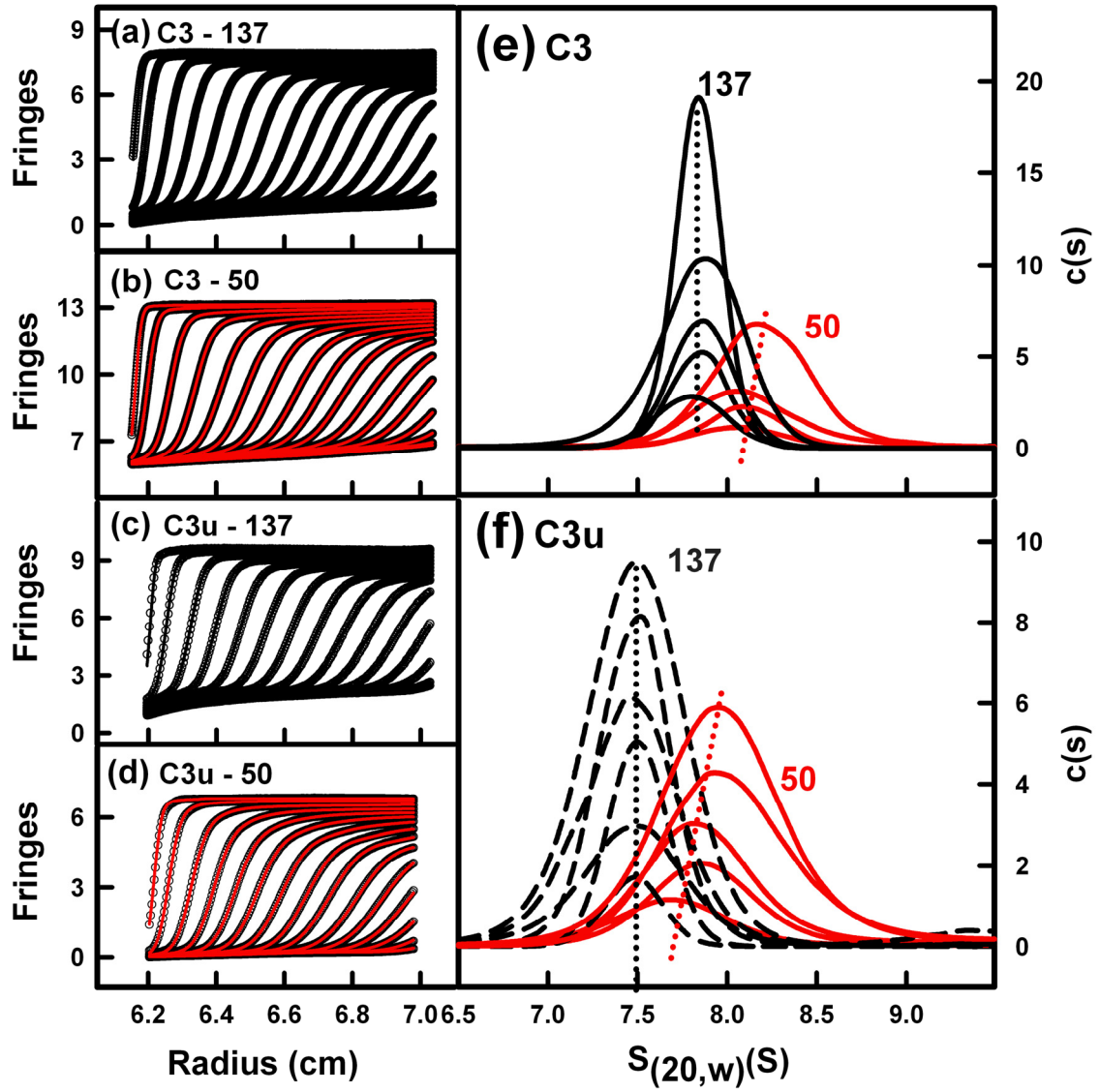


Figure 5.3 (figure legend overleaf)

Figure 5.3 Sedimentation velocity size distribution analyses *c(s)* of C3 and C3u.

(a-d) Representative interference boundary fits for C3 and C3u at about 2 mg/ml in 137 mM NaCl (black) and 50 mM NaCl (red) buffers. Only every sixth or tenth scan of the fitted boundaries are shown for clarity.

(e) The *c(s)* distributions for C3 in 137 mM (black) and 50 mM (red) NaCl buffers. In 137 mM NaCl, C3 concentrations ranged from 0.5 mg/ml to 2.2 mg/ml. In 50 mM NaCl, C3 concentrations ranged from 0.3 mg/ml to 2.3 mg/ml. Here and in (f), the dotted lines indicate the movement of $s_{20,w}$ with change of concentration.

(f) The *c(s)* distributions for C3u in 137 mM (black) and 50 mM (red) NaCl buffers. In 137 mM NaCl, C3u concentrations range from 0.3 mg/ml to 2.4 mg/ml. In 50 mM NaCl, C3u concentrations range from 0.4 mg/ml to 2.1 mg/ml.

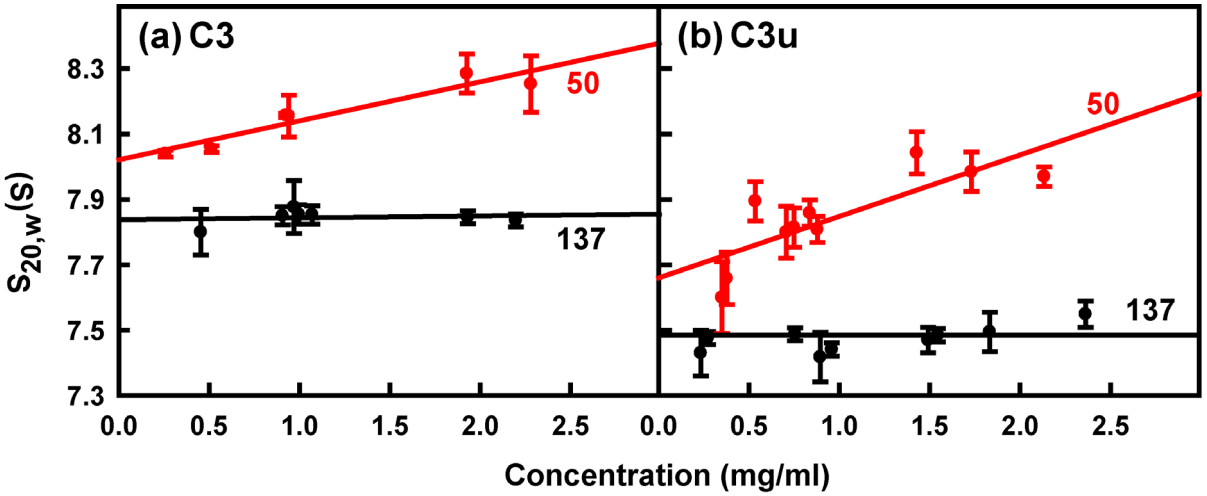


Figure 5.4 Concentration dependence of the sedimentation coefficients $s_{20,w}$ for C3 and C3u.

(a) The $s_{20,w}$ values for C3 in 50 mM (red) and 137 mM (black) were fitted by linear regression to result in $s^0_{20,w}$ values at zero concentration of 7.85 ± 0.05 S in 137 mM NaCl and 8.02 ± 0.12 S in 50 mM NaCl. Statistical error bars are shown where visible.

(b) The corresponding $s_{20,w}$ values for C3u in 50 mM (red) and 137 mM (black) were fitted by linear regression to result in $s^0_{20,w}$ values at zero concentration of 7.47 ± 0.07 S in 137 mM NaCl and 7.66 ± 0.19 S in 50 mM NaCl.

(ii) In 50 mM NaCl, the major $c(s)$ peak for C3 and C3u both showed a clear concentration dependence (Figure 5.3(e,f)). The molecular weight and f/f_o values for C3 were 190 ± 18 kDa and 1.38 ± 0.09 respectively; those for C3u were 175 ± 20 and 1.37 ± 0.12 respectively. The $s_{20,w}$ values for C3 ranged from 8.04 S to 8.25 S, while those for C3u increased by a larger proportion from 7.59 S to 8.02 S for C3u. The extrapolated $s^0_{20,w}$ values for C3 and C3u were 8.02 ± 0.12 S and 7.66 ± 0.19 S respectively, being within error of the corresponding values in 137 mM NaCl. It is of interest that these values are slightly increased in 50 mM compared to 137 mM, because this implies that both proteins became more compact in 50 mM NaCl.

The concentration dependence in 50 mM NaCl resembles that seen for the C3d fragment of C3, which gave a single $s_{20,w}$ value that was unchanged at 3.0 S in 137 mM NaCl, but gave three $c(s)$ peaks at $s_{20,w}$ values of 3.0 S, 4.3 S and 6.0 S in 50 mM NaCl (Gilbert *et al*., 2005; Li *et al*., 2008; Chapter 4). Further inspection of the $c(s)$ analyses for C3 and C3u revealed that a second minor $s_{20,w}$ peak at 11.2 ± 1.0 S and 11.5 ± 1.3 S respectively was consistently observed for both proteins (Figure 5.5(a)). The molecular weights from the $c(s)$ peaks corresponded to 330 ± 36 kDa and 305 ± 36 kDa for C3 and C3u respectively; this is less than double the sequence-predicted molecular weight of 378 kDa. This lower-than-expected molecular mass indicate that this peak is not a stable C3 or C3u dimer but is instead a reaction boundary corresponding to a rapid equilibrium between monomeric and dimeric protein on the time scale of sedimentation (Dam & Schuck, 2005). Reaction boundaries are peaks that correspond to the co-sedimentation of two different species, and exhibit well-defined sedimentation coefficients between the values of the two different species. Peak integrations showed that the estimated proportion of dimer is low at 6 ± 3% for C3 and 8 ± 2% for C3u in 137 mM NaCl, and 5 ± 3% for C3, but increases up to 23% for C3u in 50 mM NaCl (Figure 5.5(b,c)). The existence of dimer in rapid equilibrium with monomer accounts for the concentration dependences for C3 and C3u in 50 mM (Figure 5.4). Given the knowledge that C3d itself self-associates in 50 mM NaCl, the greater salt dependence seen for C3u in 50 mM NaCl compared to 137 mM NaCl indicated that the C3d region is significantly more exposed in C3u than in C3.

Previous analytical ultracentrifugation experiments reported $s^0_{20,w}$ values for C3 of 9.5 S (Bokisch *et al*., 1969), 8.3 ± 0.1 S (Molenaar *et al*., 1973), 7.3 ± 0.4 S (Paques, 1980), and 7.3 ± 0.3 S (Perkins & Sim, 1986). While the earlier work gave variable

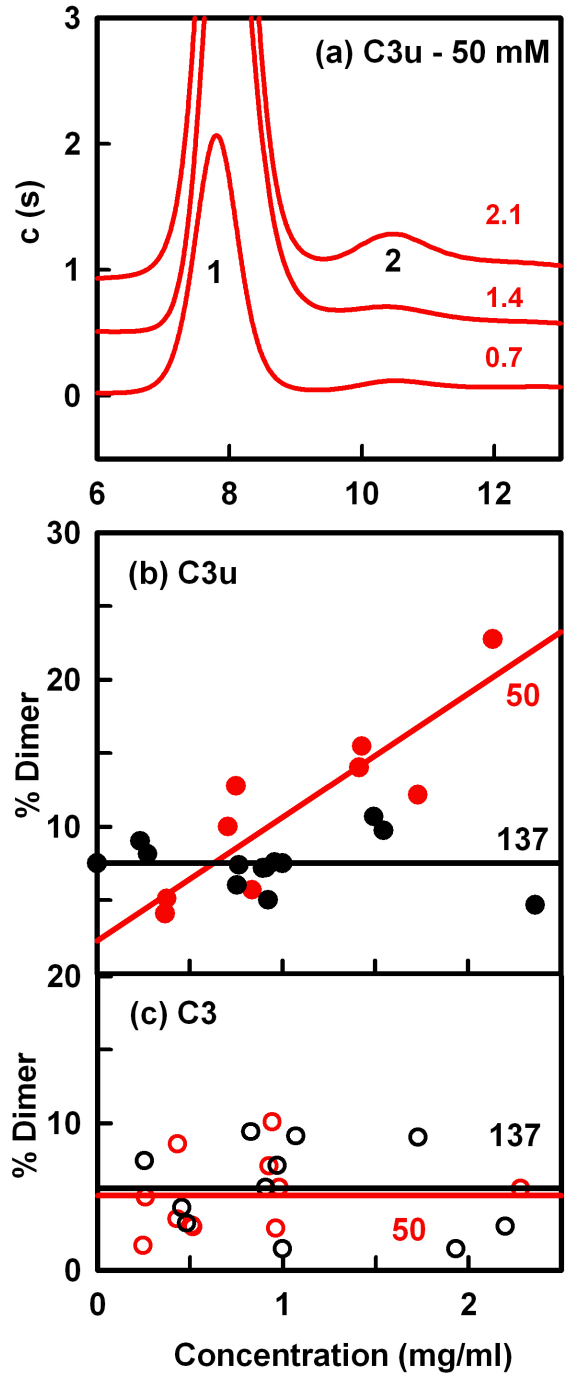


Figure 5.5 Dimer formation in C3 and C3u from size distribution analyses *c*(*s*).

(a) C3u in 50 mM NaCl buffer is shown at 0.7 mg/ml, 1.4 mg/ml and 2.1 mg/ml. The monomer peak is labelled as 1 (Figure 5.3(e,f) and the dimer peak is labelled as 2).

(b) The concentration dependence of the percentage C3u dimer formation in 50 mM NaCl (red) and 137 mM NaCl buffers (black) by linear regression.

(c) The concentration dependence of the percentage C3 dimer formation in 50 mM NaCl and 137 mM NaCl buffers. The mean values are shown by the lines.

results, their mean value is close to the C3 value of 7.85 ± 0.06 S measured in 137 mM NaCl in the present study.

(5.2.2) X-ray solution scattering data for C3 and C3u

Solution scattering is a diffraction technique that studies the overall structure of macromolecules in solution (Perkins *et al*., 2008, 2009). Freshly prepared C3 and C3u were studied in 50 mM and 137 mM NaCl buffer. The domain arrangement and self-association of C3 and C3u were assessed by Guinier analyses of the X-ray $I(Q)$ data in two different Q ranges ($Q = 4\ \pi \sin\theta / \lambda$; 2θ = scattering angle; λ = wavelength). The overall radius of gyration R_G monitors the degree of elongation of C3 and C3u, while that for its cross-section R_{XS} monitors the elongation of its cross-section shape along its longest axis. The $I(0)/c$ value is proportional to the molecular mass. All four sets of Guinier R_G analyses showed linear fits within satisfactory $Q.R_G$ and $Q.R_{XS}$ limits (Figure 5.6). The concentration dependence of the R_G values again reflected the buffer in use:

(i) In 137 mM NaCl, the R_G values for C3 and C3u were unchanged at 4.52 ± 0.08 nm and 4.88 ± 0.23 nm with protein concentration (Figure 5.7(a,c)). The larger value for C3u reflected its more elongated structure, in agreement with the $s^0_{20,w}$ values. The $I(0)/c$ values were similar at 0.0186 ± 0.0009 and 0.0196 ± 0.0013 (Figure 5.7(b,d)), showing that both C3 and C3u were similar in mass as expected. The R_{XS} values were the same at 2.54 ± 0.06 nm and 2.55 ± 0.16 nm for C3 and C3u (Figure 5.7(e,f)).

(ii) In 50 mM NaCl, the R_G values for C3 and C3u increased linearly with concentration. At zero concentration, the R_G values were 4.87 ± 0.26 nm and 5.16 ± 0.22 nm respectively. That these values are slightly larger than those at 137 mM is attributed to the increased proportion of dimeric species in reduced salt conditions, meaning that the mean R_G value corresponds to an average of monomer and a small amount of dimer. The $I(0)/c$ values were similar at 0.0199 ± 0.0016 and 0.0178 ± 0.0011, and agree with those at 137 mM NaCl when extrapolated to zero concentration. This shows that the $I(0)/c$ values correspond to similar masses and that the monomer-dimer equilibrium is reversible. The R_{XS} values also became the same at 2.50 ± 0.17 nm and 2.56 ± 0.03 nm for C3 and C3u when extrapolated to zero concentration.

The distance distribution function $P(r)$ leads to the determination of overall lengths L following an assumption of the value of the maximum dimension D_{max}. The R_G values calculated from the $P(r)$ analyses agree with those from the Guinier analyses (Table 5.1). In 137 mM NaCl, the length L of C3 and C3u in 137 mM NaCl is 16 nm

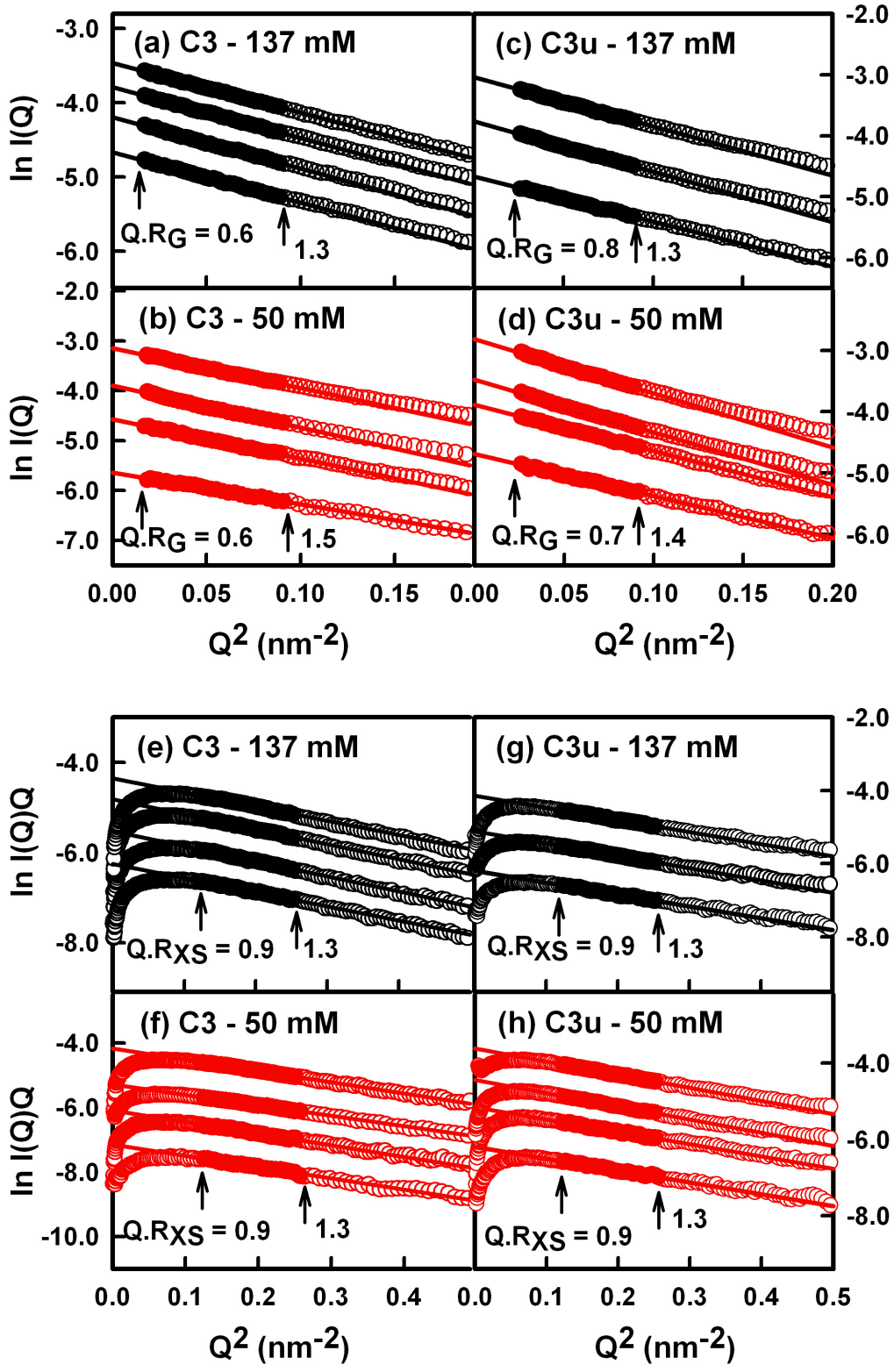


Figure 5.6 Experimental X-ray Guinier R_G and cross-section R_{XS} analyses of C3 and C3u. The Q ranges used for the R_G fits were 0.13 to 0.30 nm^{-1} for C3, and 0.16 to 0.30 nm^{-1} for C3u. . The Q ranges used for the R_{XS} fits were 0.35 to 0.50 nm^{-1} for C3 and C3u. The data points are denoted by open circles; the filled circles were used to determine the R_G (a-d) and R_{XS} (e-h) values based on the best fit lines as shown.

(a,e) C3 concentrations between 0.5 mg/ml to 1.2 mg/ml in 137 mM NaCl (black) from bottom to top.

(b,f) C3 concentrations between 0.2 mg/ml to 2.0 mg/ml in 50 mM NaCl (red).

(c,g) C3u concentrations between 0.5 mg/ml to 1.1 mg/ml in 137 mM NaCl.

(d,h) C3u concentrations between 0.5 mg/ml to 1.7 mg/ml in 50 mM NaCl.

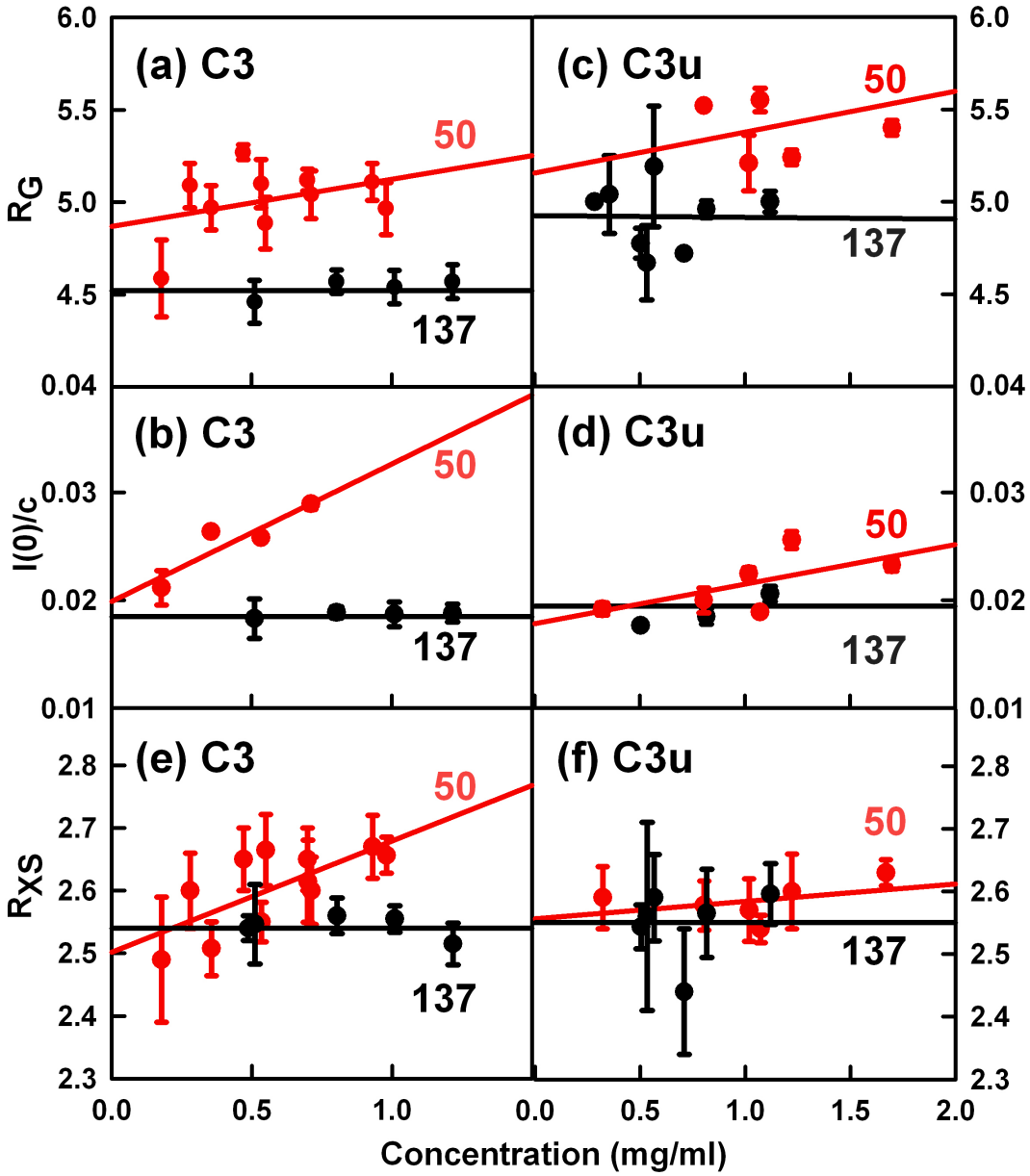


Figure 5.7 Concentration dependence of the Guinier R_G, $I(0)/c$ and R_{XS} values for C3 and C3u. Each value was measured in quadruplicate and averaged for linear regression fits. Statistical error bars are shown where visible.

(a, c) The R_G values are shown for 137 mM NaCl (black) and 50 mM NaCl buffer (red).

(b, d) The corresponding $I(0)/c$ values are shown for 137 mM and 50 mM NaCl buffer. The fitted line shown for 137 mM NaCl buffer is the mean value.

(e, f) The R_{XS} values are shown for 137 mM NaCl and 50 mM NaCl buffer.

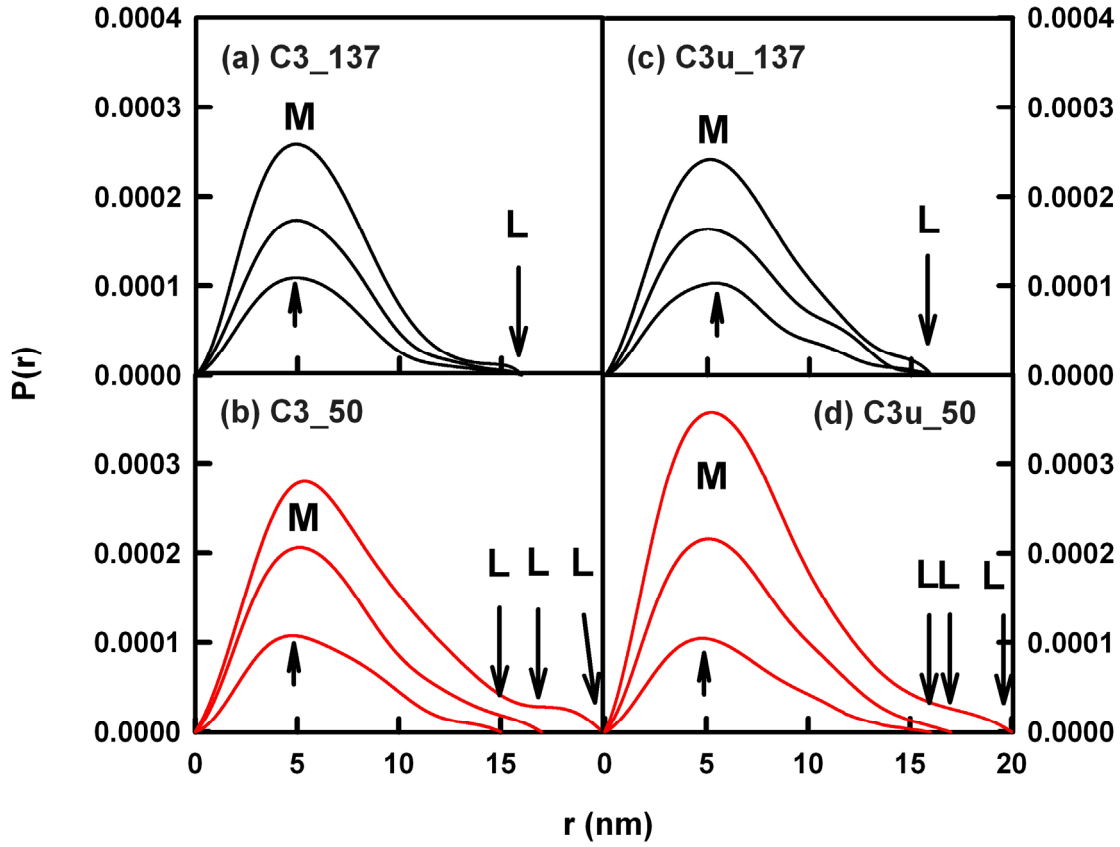


Figure 5.8 X-ray distance distribution function *P*(*r*) analyses for C3 and C3u. For C3 and C3u in 137 mM NaCl, the peak maximum *M* is observed at 5.0 nm. For C3 and C3u in 50 mM NaCl, M is observed at 5.0 to 5.4 nm for C3 and 5.0 to 5.2 for C3u.

(a) C3 between 0.5 mg/ml to 1.2 mg/ml in 137 mM NaCl buffer showed an unchanged maximum length *L* of 16 nm.

(b) C3 in 50 mM NaCl buffer showed *L* values of 15 nm (0.6 mg/ml), 17 nm (0.7 mg/ml) and 20 nm (0.9 mg/ml).

(c) C3u between 0.5 mg/ml to 1.1 mg/ml in 137 mM NaCl buffer showed an unchanged *L* value of 16 nm.

(d) C3u in 50 mM NaCl buffer showed L values of 16 nm (0.5 mg/ml), 17 nm (1.0 mg/ml) and 20 nm (1.7 mg/ml).

Table 5.1 X-ray scattering and sedimentation coefficient values from the C3 and C3u experiments.

	R_G (nm) [a]	R_{XS} (nm)	D_{max} (nm) [b]	$s^0_{20,w}$ (S)
Experimental				
C3 (50 mM NaCl)	4.87 ± 0.26 4.80 ± 0.18	2.50 ± 0.17	15-20	8.02 ± 0.12
C3 (137 mM NaCl)	4.52 ± 0.08 4.67 ± 0.19	2.54 ± 0.04	16	7.85 ± 0.05
C3u (50 mM NaCl)	5.16 ± 0.22 5.35 ± 0.16	2.56 ± 0.03	16-20	7.66 ± 0.19
C3u (137 mM NaCl)	4.88 ± 0.23 5.07 ± 0.21	2.55 ± 0.10	16	7.44 ± 0.07

[a] The experimental values are extrapolated to zero concentration. The first value is from the Guinier R_G analyses (Figure 5.6 and 5.7) and the second value is from the GNOM *P(r)* analyses (Figure 5.8).

[b] The D_{max} values in here are from *P(r)* analyses, while the values in Table 5.2 and Table 5.3 are calculated using HYDROPRO.

(Figure 5.8(a,c)). In 50 mM NaCl, the L values increase from 15-16 nm at the lowest concentrations of 0.5-0.6 mg/ml up to 20 nm at 1.0-1.7 mg/ml. This concentration dependence is consistent with Figure 5.7.

Previous X-ray scattering experiments were performed in different buffers (50 mM potassium phosphate, 100 mM KC1, 5 mM EDTA, pH 7.0, or 12.4 mM sodium phosphate, 200 mM NaC1, 0.5 mM EDTA, pH 7.0) (Perkins & Sim, 1986). The previous R_G values for C3 in 200 mM NaCl (Perkins & Sim, 1986) were 5.2 ± 0.1 nm, which is larger than the present C3 R_G values (Table 5.1). The difference is attributable to improvements in sample purifications in the present study, most notably the avoidance of protein aggregation by gel filtration immediately prior to beam time, and instrumental improvements at the ESRF compared to those at DESY Hamburg and SRS Daresbury, most notably in terms of better detector sensitivities, and a larger Q range. The previous R_{XS} value for C3 in 200 mM NaCl (Perkins & Sim, 1986) of 2.6 ± 0.1 nm agrees with the current C3 R_{XS} value of 2.54 nm (Table 5.1).

(5.2.3) Constrained modelling of the C3 and C3u solution structures

C3 is converted to C3b by the removal of its anaphylatoxin C3a domain. Comparatively large conformational changes in several MG domains occur during the conversion of inactive C3 to active C3b. C3b has 12 domains, namely the MG1-MG8, LNK, and C345C domains within the C3c region and the CUB and TED domains outside this region (Figure 5.9(a)). The CUB and TED domains are connected by linker peptides of length seven to ten residues to the MG7 and MG8 domains, denoted by 1-4 in Figure 5.9(a). These peptides are significantly altered in conformation during the large movement of the CUB and TED domains from their location proximate to the C345C domain in C3 to their location adjacent to the MG1 domain in four C3b crystal structures.

The crystal structures were tested for their agreement with the sedimentation data. The C3 crystal structure gave a calculated $s^0_{20,w}$ value of 8.11 S (Table 5.2), which agreed well with the experimental values of 8.02 S and 7.85 S in two buffers (Table 5.1), given that the accuracy of the method is ± 0.21 S (Perkins *et al*., 2009). The four similar C3b crystal structures gave calculated $s^0_{20,w}$ values in a range of 7.51-7.69 S, which agrees well with the experimental value of 7.66 S for C3u in 50 mM NaCl. Interestingly, the deviant C3b crystal structure gave a calculated $s^0_{20,w}$ value of 7.37 S, which agrees

Table 5.2 Single models fit for the solution structures of C3 and C3u.

	Models	Hydrated spheres N	R_G (nm) [a]	R_{XS} (nm)	D_{max} (nm) [b]	R-factor (%)	$s^0_{20,w}$ (S)
Crystal structure							
C3 (PDB 2a73)	1	1863	4.35	2.45	16.4	3.1	8.11
C3b (PDB 2i07)	1	1845	4.64	2.48	17.7	4.8	7.69
C3b (PDB 2icf)	1	1895	4.58	2.52	17.5	4.5	7.56
C3b (PDB 2win)	4	1804-1856	4.59-4.62	2.46-2.50	17.3-17.6	4.6-5.1	7.51-7.59
C3b (PDB 2wii)	1	1856	4.50	2.56	16.0	4.5	7.62

[a] The R_G values are calculated from individual models by Debye equation adapted to spheres using the same Q range as the X-ray scattering experiments.

[b] The values in here are calculated using HYDROPRO.

well with the experimental value of 7.47 S for C3u in 137 mM NaCl. It is concluded that, if the TED/CUB domains become extended away from the eight MG domains in C3b in 137 mM NaCl, this movement will account for the difference in $s^0_{20,w}$ values seen between 50 mM and 137 mM NaCl.

The crystal structures were next tested for their agreement with the scattering data. The experimental R_G values of 4.52 - 4.87 nm for C3 were similar to the calculated R_G value of 4.35 nm for C3 within error (Table 5.1, Table 5.2). The experimental R_G values of 4.88 - 5.16 nm for C3u were also similar to the calculated R_G values of 4.50-4.64 nm for the four C3b crystal structures within error. That the experimental values were slightly larger than the calculated values is attributed to the sensitivity of the scattering experiment to small amounts of dimers. A good curve fit was obtained between the C3 crystal structure and C3 in 137 mM NaCl with a low R-factor of 3.1%. The four similar C3b crystal structures gave larger R-factors of 4.5-5.1% with C3u in 137 mM NaCl. An improved R-factor of 3.9% was obtained with the deviant C3b crystal structure. This showed again that, if the TED/CUB domains became extended away from the MG domains, slightly improved agreement between the solution and crystal structures was obtained.

The location of the CUB and TED domains is crucial to the biological function of C3, C3u and C3b. Solution structures for C3 and C3u in 137 mM NaCl were determined by constrained scattering modelling, in which the C3 and C3b crystal structures were used as the starting constraints (Section 5.4.5). The scattering modelling of C3u assumed that the functional similarity between C3u and C3b meant that the C3b crystal structure (PDB code 2I07) should be used to model this. A range of different searches was used to show that the C3u solution structure did not depend on the modelling strategy in use. To determine the CUB/TED location in the solution structure of C3u, the four linkers were each individually conformationally randomised in four searches in order to create up to 8,000 different trial structures per search in which the CUB and TED domains were distributed in all orientations relative to the MG1-MG8 domains. That a sufficiently wide range of C3u conformations was generated in each search by the modelling procedure is shown by the superimposition of the first 100 models created from the linker 2 randomisation (Figure 5.10). The side and top views show a full degree of conformational variability in the position of the CUB-TED domains (blue). The best fit model from the linker 2 search (green circle from Figure 5.9c) is shown by the pink ribbon denoting the

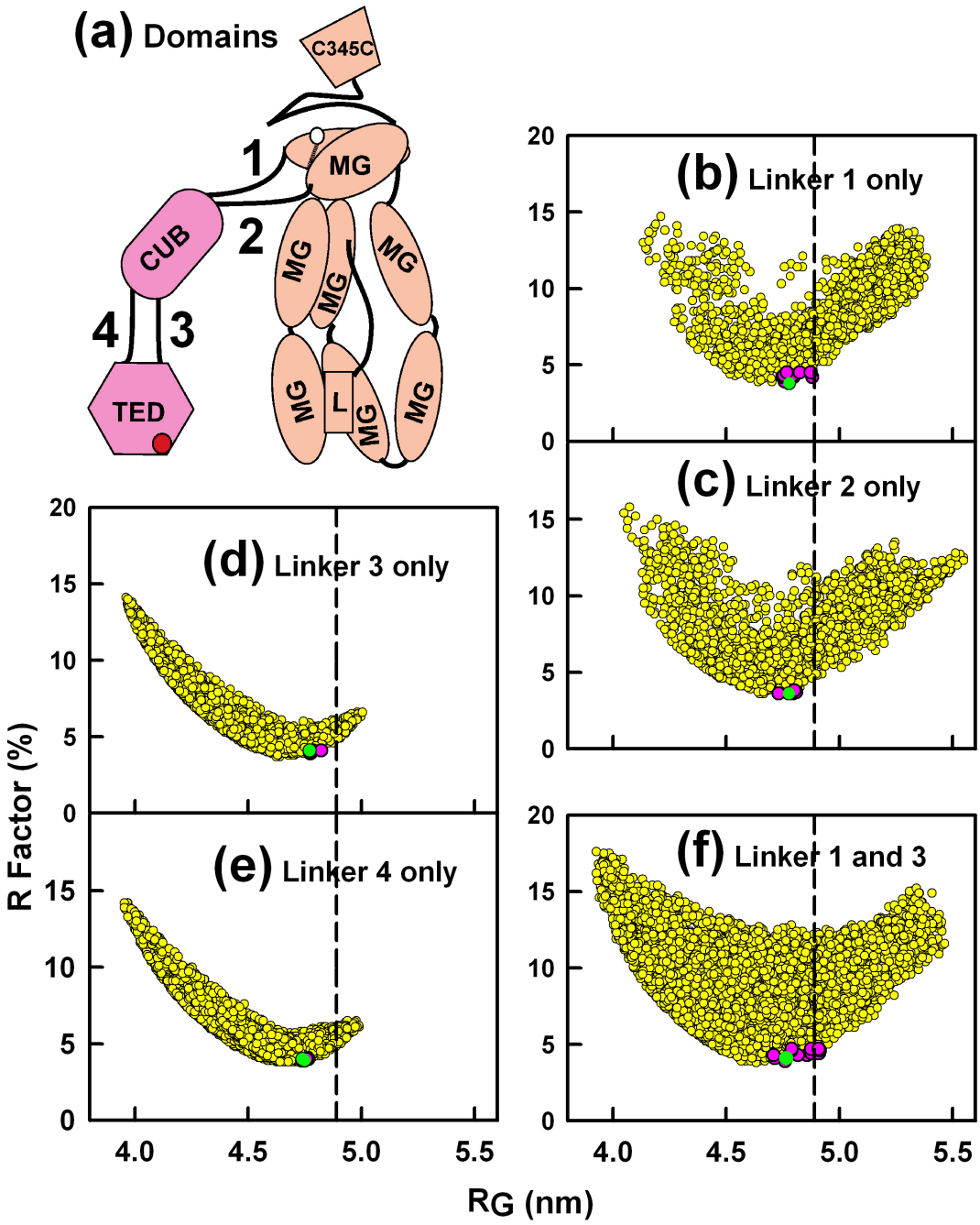


Figure 5.9 (figure legend overleaf)

Figure 5.9 Constrained modelling analyses of the C3u solution structure.

(a) Cartoon of the 12 domains in C3b, with the CUB and TED domains shown in pink, and the other domains (MG1- MG8, C345C and L) comprising the C3c segment of C3b shown in tangerine. The interdomain linkers varied in the modelling are labelled as 1, 2, 3 and 4.

(b) The *R* factor values for 6,000 trial models of C3u in which linker 1 was varied are compared with the corresponding R_G values. The vertical dashed line corresponds to the experimental R_G value. The 13 good fit models are shown by pink circles, of which the best-fit model with the lowest R factor is denoted by the green circle.

(c) The linker 2 search was based on 4,650 trial models. The five good fit models and single best fit best-fit model are denoted by pink and green circles respectively.

(d) The linker 3 search was based on 6,000 models. The six good fit models and single best fit best-fit model are denoted by pink and green circles respectively.

(e) The linker 4 search was based on 4,200 models. The 15 good fit models and single best fit best-fit model are denoted by pink and green circles respectively.

(f) The linker 1 and 3 search was based on 8,000 models. The 13 good fit models and single best fit model are denoted by pink and green circles respectively.

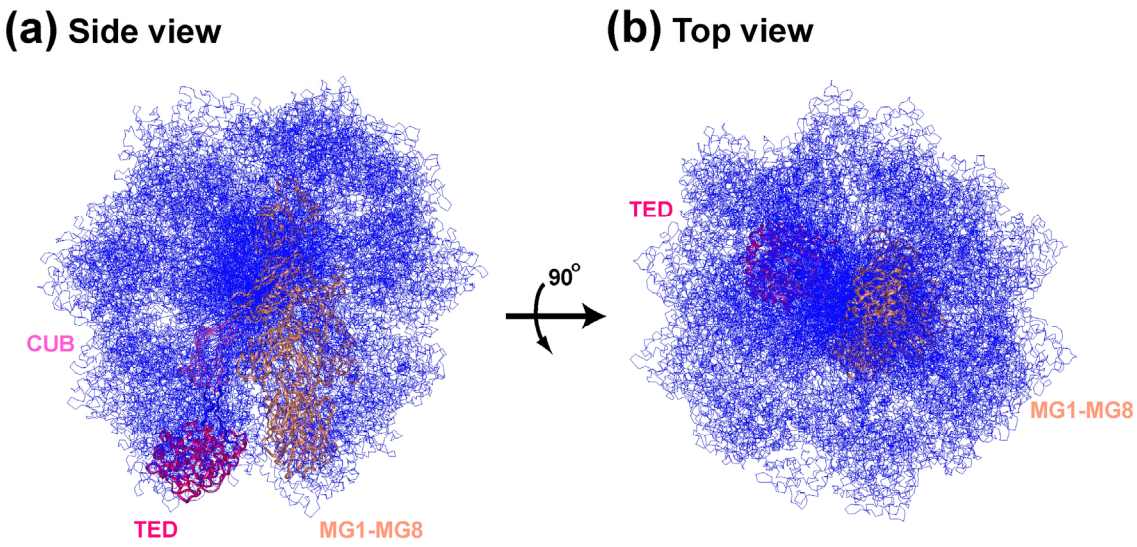


Figure 5.10 Superimposition of 100 randomised C3u models created from the linker 2 search. The models were superimposed on the invariant C3c segment of C3b shown in tangerine (PDB code 2I07). The best fit TED-CUB model is highlighted as a pink ribbon. The positions of the TED-CUB domains from the first 100 models from the 4,650 used in this search are shown in blue. The views in (a) and (b) correspond to the side and top views in which the side view is rotated by 90° as arrowed.

CUB and TED domains, with the remaining domains (MG1- MG8, C345C, LNK) shown in tangerine.

The predicted X-ray curves calculated from the models after conversion to Debye spheres in the five searches of Figure 5.9 were compared with the experimental C3u curve in PBS 137 mM NaCl buffer. The graphs of *R*-factor (goodness-of-fit) vs. the R_G values showed that the *R*-factors ranged between 3.6% to 17.6% and the R_G values ranged from 3.8 nm to 5.6 nm (Figure 5.9(b-f)). The experimental R_G value of 4.97 nm is close to the minimum *R*-factor values. By applying filters based on the total number of Debye spheres (a monitor of whether the CUB and TED domains sterically overlap with the MG domains), and the requirement that the R_G and R_{XS} values were within ± 5 % of the experimental values, and sorting the output in order of lowest *R*-factors, 5-15 best fit C3u models were identified from each conformational search (pink circles in Figure 5.9(b-f), of which the best-fit model is denoted by the green circle. The best fit C3u model with an *R*-factor of 3.6% was determined from the Linker 2 search. Comparison of the calculated and experimental C3u $I(Q)$ curves showed good visual agreement (Figure 5.11(b)). Likewise the modelled $P(r)$ curve showed a major peak at r = 5.45 nm and L of 16 nm, indicating good agreement with Figure 5.8(b).

The best-fit C3u models from the four Linker 1-4 searches consistently showed that the TED domain was well separated from the MG1 domain, unlike in the C3b crystal structures (Figure 5.12). The ribbon models in green correspond to the best fit with the lowest R-factor. The different types of searches meant that the R_G values showed a reduced range of values in Figures 5.9(d,e) compared to Figures 5.9(b,c). The randomisation of either Linker 1 or 2 moved the rigid CUB-TED domain pair into one set of similar best-fit locations, while the randomisation of either Linker 3 or 4 moved the TED domain away from the MG domains. The overall best-fit C3u models with the lowest R-factors resulted from the Linker 2 search. There, the separation of 6.0 ± 0.9 nm between the TED and the MG1 domains in the best-fit C3u models is significantly wider than that of 3.1 ± 0.1 nm seen in the other four crystal structures for C3b (PDB codes 2I07, 2ICF, 2WIN, 2WII) (Janssen *et al*., 2006; Wiesmann *et al*., 2006; Wu *et al*., 2009; Rooijakkers *et al*., 2009). The deviant C3b crystal structure (PDB code 2HR0) (Ajees *et al*., 2006) showed a separation of 7.1 nm, this being more similar to our best-fit C3u structure.

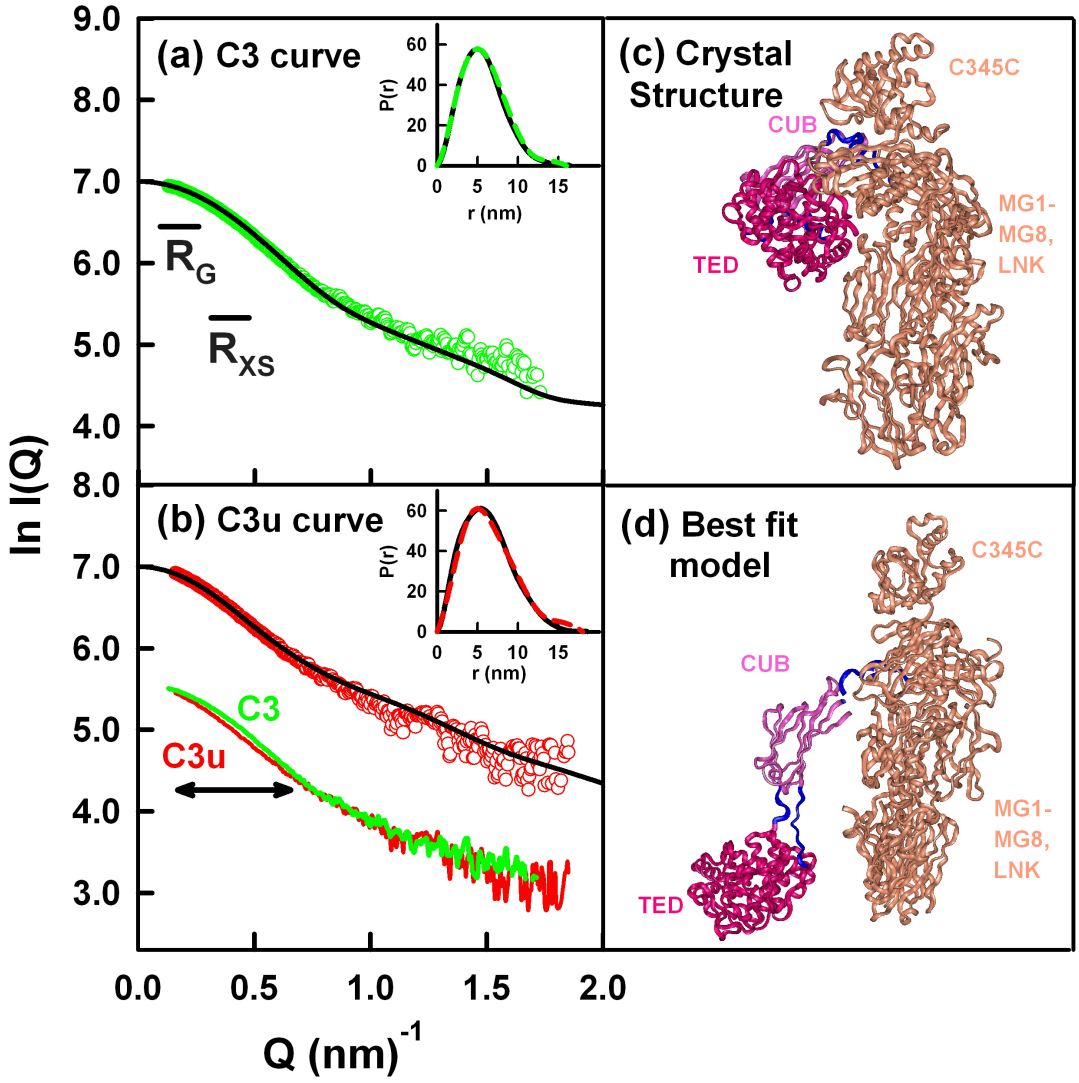


Figure 5.11 (figure legend overleaf).

Figure 5.11 X-ray scattering curve fits for C3 and C3u. The experimental data for C3 and C3u in 137 mM NaCl buffer correspond to 1.2 mg/ml and 1.1 mg/ml respectively.

(a) The experimental $I(Q)$ and $P(r)$ curves for C3 are shown in green. The modelled curves based on the C3 crystal structure (PDB code 2A73) are shown in black.

(b) The experimental $I(Q)$ and $P(r)$ curves for C3u are shown in red. The curves for the best-fit C3u model are shown in black. The experimental $I(Q)$ curves for C3 (green) and C3u (red) are also shown superimposed upon each other, where the main difference between the curves is highlighted by a black bar

(c) A ribbon view of the C3 crystal structure, with the TED and CUB domains highlighted in red. Here and in (d), the interdomain linker peptides are highlighted in blue.

(d) A ribbon view of the C3u best fit model, with the TED and CUB domains highlighted in red.

Table 5.3 X-ray scattering and sedimentation coefficient modelling fits for the solution structures of C3 and C3u.

	Filter	Models	Hydrated spheres N	R_G (nm)	R_{XS} (nm)	D_{max} (nm) [a]	R-factor (%)	$s^0_{20,w}$ (S)
C3u modelling								
Linker 1 search	None	6000	1553-1895	4.15-5.38	1.67-2.83	n. a.	3.8-14.7	n. a.
	N, R_G, R_{XS} and R factor	13	1853-1877	4.75-4.88	2.56-2.65	17.1-18.2	3.8-4.5	7.27-7.53
	Best fit	1	1867	4.78	2.59	18.6	3.8	7.46
Linker 2 search	None	4650	1526-1902	4.08-5.55	1.66-2.90	n. a.	3.6-15.8	n. a.
	N, R_G, R_{XS} and R factor	5	1870-1886	4.73-4.81	2.55-2.68	18.0-19.0	3.6-3.8	7.33-7.53
	Best fit	1	1877	4.78	2.55	19.0	3.6	7.48
Linker 3 search	None	6000	1622-1894	3.97-5.01	1.97-2.85	n. a.	3.7-14.1	n. a.
	N, R_G, R_{XS} and R factor	6	1864-1883	4.77-4.82	2.54-2.58	17.9-18.2	3.9-4.1	7.33-7.46
	Best fit	1	1871	4.77	2.55	17.9	4.1	7.44
Linker 4 search	None	4200	1623-1881	3.96-4.82	1.97-2.78	n. a.	3.8-14.2	n. a.
	N, R_G, R_{XS} and R factor	15	1851-1873	4.74-4.77	2.55-2.67	17.3-18.4	3.9-4.0	7.33-7.59
	Best fit	1	1860	4.74	2.59	18.0	3.9	7.46
Linker 1 and 3 search	None	8000	1453-1890	3.93-5.45	1.52-2.90	n. a.	3.8-17.6	n. a.
	N, R_G, R_{XS} and R factor	13	1853-1888	4.71-4.91	2.50-2.69	16.1-18.9	3.9-4.7	7.29-7.55
	Best fit	1	1878	4.76	2.50	18.9	4.1	7.48

n. a. Not available.

[a] The values in here are calculated using HYDROPRO.

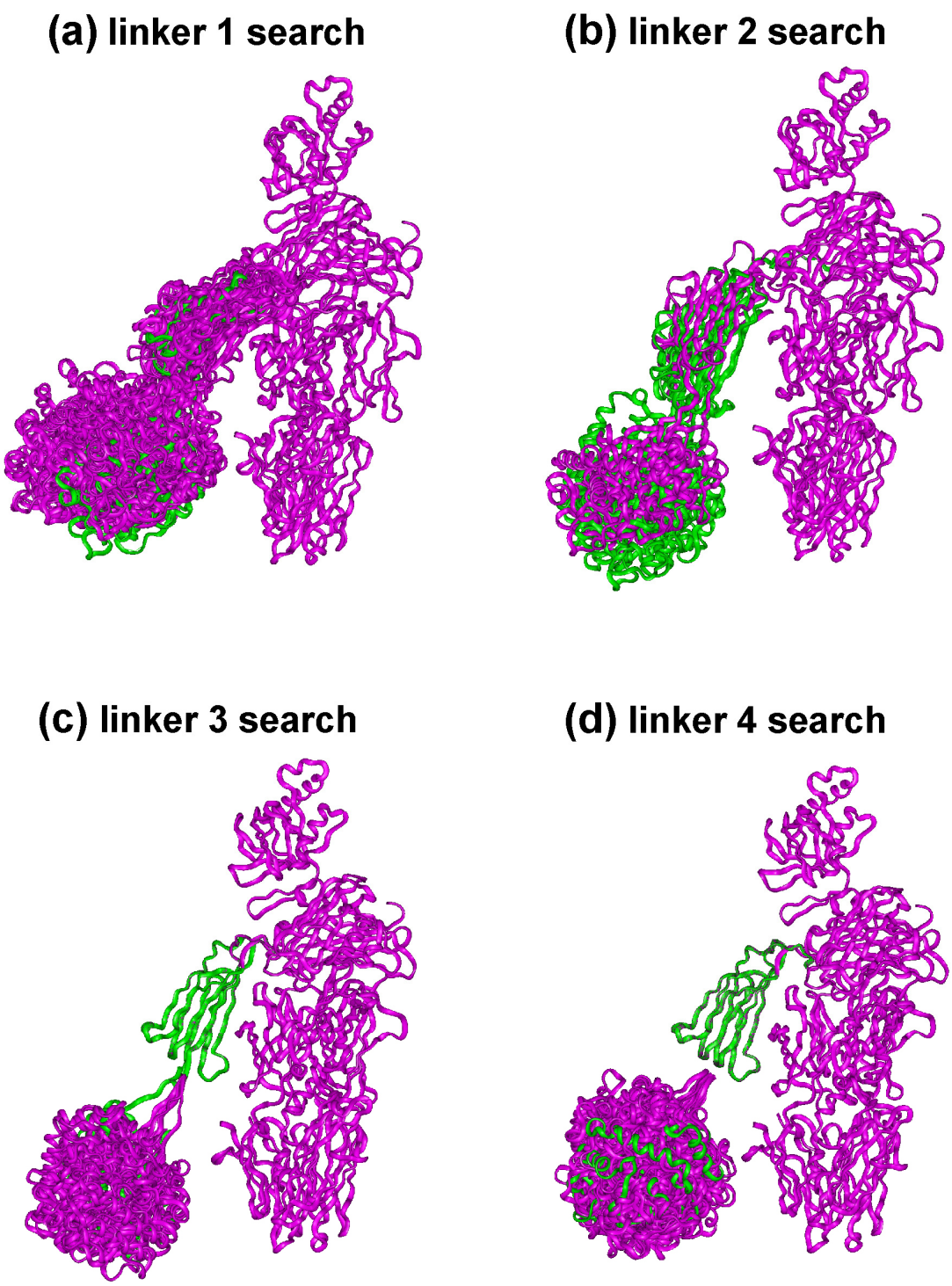


Figure 5.12 Comparisons of the best fit models for C3u generated from each of the linker 1-4 searches. The best-fit model with the lowest R-factor is shown in green, and the remaining 5-15 good fit models are shown in pink. The TED-CUB and MG1-MG8 domains are in pink.

As one control of the modelling procedure, the $I(Q)$ and $P(r)$ curves calculated from the C3 crystal structure were compared with the experimental C3 curves in 137 mM NaCl at 1.2 mg/ml (Figure 5.11a). A good visual match was observed, in which the major peak at r = 5.0 nm in the crystallographic $P(r)$ curve reproduced the experimental $P(r)$ curve of Figure 5.8a. It was concluded that the crystal and solution structures of C3 are similar. As a second control, the significance of the structural difference between C3 and C3u is shown by the comparison of their X-ray scattering curves, where clear intensity differences are seen in the Q range between 0.1-0.6 nm^{-1} (Figure 5.11b) as well as the neutron scattering curves (Figure 5.13). Models were created from a new double Linker 1 and 3 search performed using the experimental curve of C3 in PBS 137 mM NaCl buffer with 8,000 trial C3b models. In Figure 5.14, the resulting R-factor *vs.* R_G analysis for C3 differed significantly from the same double Linker 1 and 3 search performed for C3u. The minimum is shifted to lower R_G values, and interestingly showed lower R-factors than the C3u analysis. The superimposition of the 15 best-fit C3b structures using the fit with the C3 scattering curve showed that seven of these resembled the C3 crystal structure itself, while the other eight showed a broad range of other conformations. Even though this search based on a C3b crystal structure had not allowed for the significant shifts of the MG domains between C3b and C3, this fit outcome for C3 validated the modelling procedure for C3u.

(5.3) Discussion

C3, C3u and C3b are important for the initiation, amplification and effector functions of complement (Walport, 2001; Law & Reid, 1995). The "tick-over" conversion of C3 to C3u leads to alternative pathway activation, because C3u exhibits C3b-like function and forms an initial convertase with the Bb fragment of Factor B to enable the rapid generation of active C3b. We have determined solution structures for C3 and C3u by constrained scattering modelling. The high quality of the curve fits are similar to those obtained for other complement proteins and antibodies (Furtado *et al*., 2008; Bonner *et al*., 2009). While the C3 solution structure corresponded well with its crystal structure, the C3u solution structure is distinct to that of four similar C3b crystal structures (Figures 5.15(a,b)). The TED domain in C3u is significantly extended away from the MG1-MG8 domains in solution, and is not proximate to the MG1 domain. This is different from the four C3b crystal structures that show that the TED and MG1 domains are in contact with each other. A small variability in the four C3b crystal structures suggests that the TED and MG1 domains do not form a precise interaction. A

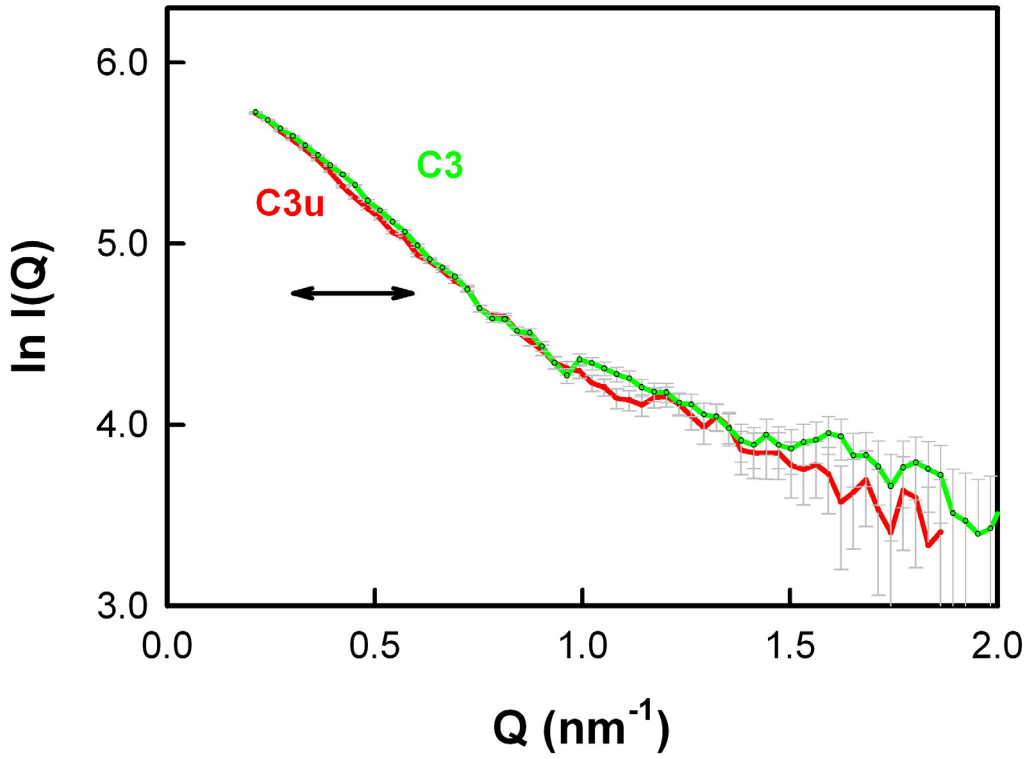


Figure 5.13 Neutron scattering curve for C3 and C3u. This was recorded on instrument SANS2D at ISIS TS2. The experimental data for C3 (green) and C3u (red) in 137 mM NaCl, $^{2}H_2O$ PBS buffer corresponding to 2.1 mg/ml and 1.7 mg/ml respectively. The experimental $I(Q)$ curves for C3 (green) and C3u (red) in neutron are shown superimposed upon each other, where the main difference between the curves is highlighted by a black bar.

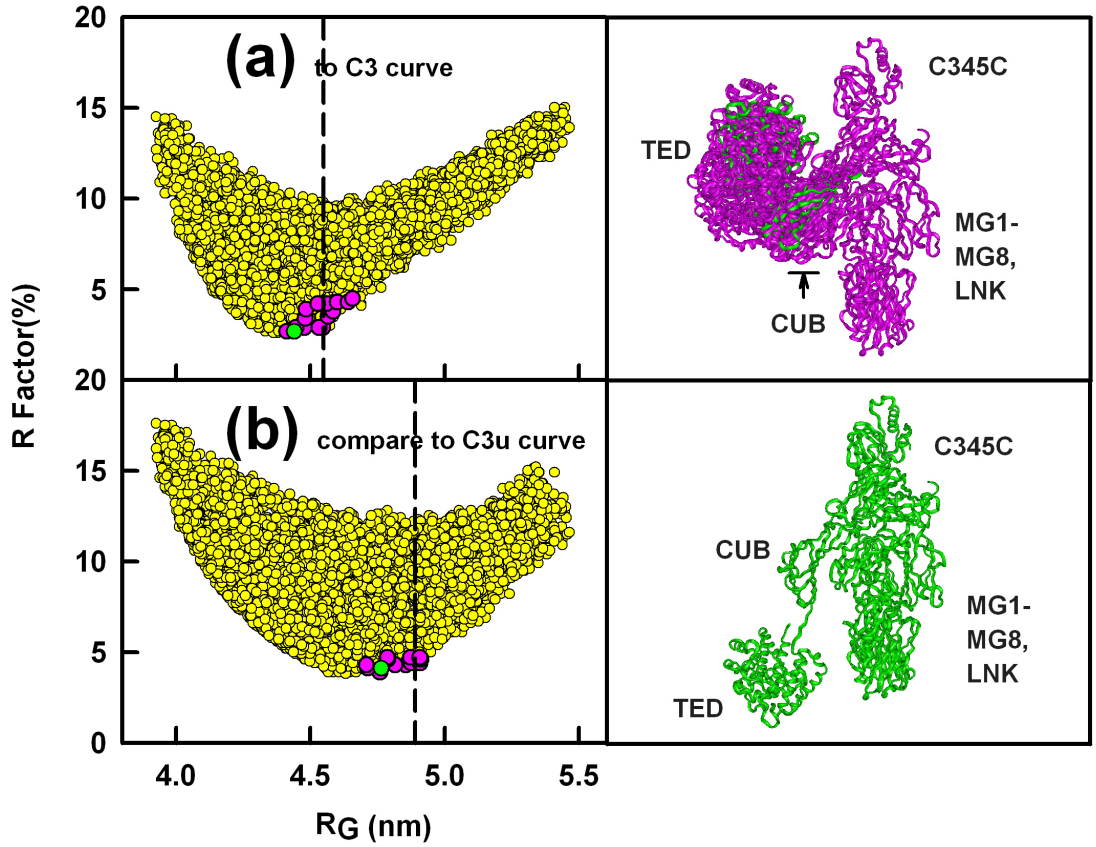


Figure 5.14 Comparison of the constrained modelling fits for the Linker 1 and 3 search with 8,000 models with the experimental *I*(*Q*) curves for C3 and C3u. The vertical dashed lines indicate the experimental R_G value for C3 in (a) and C3u in (b) in 137 mM NaCl.

(a) The 8,000 models were compared against the experimental C3 curve, giving 15 good-fit models. Seven of the 15 good-fit models resemble the C3 crystal structure, and these seven are shown in pink at the right, with the best-fit model shown in green.

(b) The 8,000 models were compared against the experimental C3u curve; this is identical with Figure 5.9(f). The best-fit model is shown as a green ribbon structure at the right.

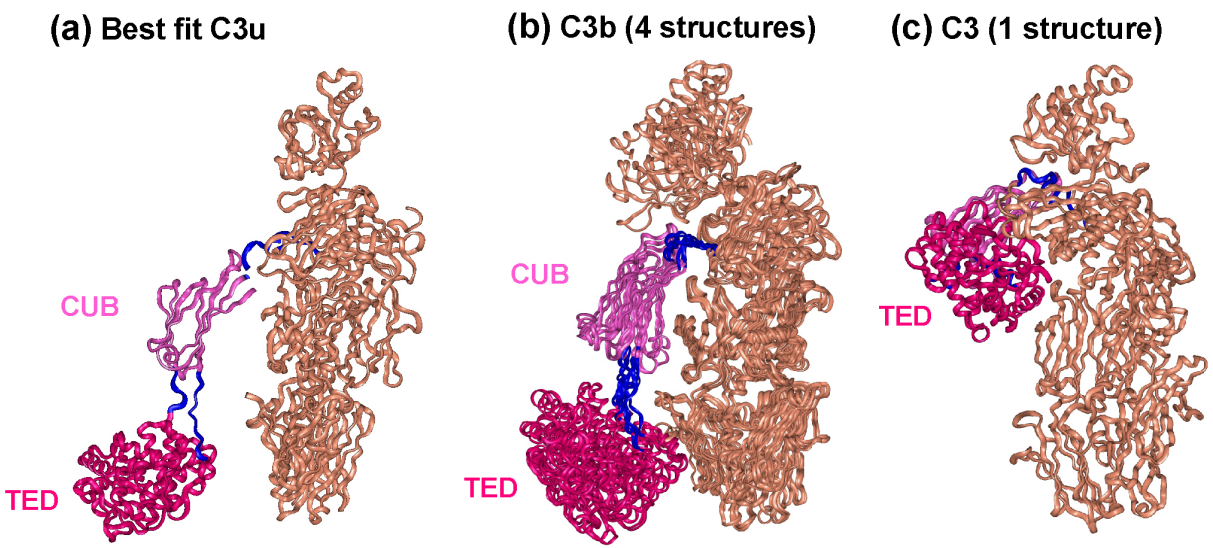


Figure 5.15 Comparison of ribbon views of the solution and crystal structures for C3u, C3b and C3. The TED and CUB domains are shown in red and pink respectively, the linkers are shown in blue, and the remaining C3c segment is tangerine.

(a) The best-fit C3u solution structure model;

(b) Superimposition of four C3b crystal structures (PDB codes 2I07, 2ICF, 2WII and 2WIN);

(c) The C3 crystal structure (PDB code 2A73).

weak interaction between the TED and MG1 domains would explain why the C3b crystal and C3u solution structures differ. Electron microscopy images are viewed in two-dimensions in comparatively harsh conditions. While shape differences between C3 and C3u were detected by electron microscopy (Nishida *et al*., 2006), their structures could not be defined in more detail. Hydrogen/deuterium exchange studies also revealed differences between C3 and C3u, and showed that C3u possessed a more open structure (Winters *et al*., 2005). Our results indicate that the arrangement of the TED and CUB domains relative to the eight MG domains exhibit previously unrecognised conformational flexibility.

Mechanistic insights are obtained by this study. Our results show that C3u represents a structure that is intermediate between those of C3 and C3b. Functionally, the ability of the TED domain to move independently from the MG domains in the C3 proteins will facilitate the attachment of C3b to an antigenic surface in a broad range of conformations independent of the remaining C3b structure. The four similar C3b crystal structures do not show that the TED domain is optimally positioned for antigen interactions, because the MG domains can potentially block steric access to one half of this. Versatile TED arrangements in the C3 proteins would permit a more efficient innate immune response by C3 and also its complement C4 and C5 homologues. Functionally also, the mobility of the TED domain relative to the fixed MG domains clarifies the need for Factor H to act as a cofactor for the Factor I-mediated cleavage of C3b. The Factor H SCR-4 domain is seen to link together the MG1 and TED domains in the crystal structure (Wu *et al*., 2009). It is deduced that Factor H restricts what would otherwise be a high conformational variability of the TED domain in order to provide an appropriate rigid template at the CUB domain for Factor I-mediated cleavage of C3.

The existence of both C3 and C3u monomer and dimer in solution was previously unrecognised. The C3 concentration in serum ranges between 1 and 1.5 mg/ml (5.3 μM to 7.9 μM) in 137 mM NaCl (Law & Reid, 1995). Given that C3 is an acute-phase reactant, its concentration is increased during inflammation, and any localised production could also result in higher C3 levels at inflammatory sites. By size-exclusion chromatography, no dimer peak was visible, however both C3 and C3u eluted earlier in 50 mM NaCl than in 137 mM NaCl buffer (Figure 5.2). Our results explain this in terms of reversible greater self-association in lower salt from our $c(s)$ size-distribution analyses and X-ray scattering data, and explains previous similar data for

C3 (Nishida *et al*., 2006). While our analysis of reversible systems has been achieved using sedimention boundary analysis, it is not rigorous. Even though there are good indications that dimerisation is occurring and differs with salt, the possible effects of conformational change, different hydration levels and non-ideality may require consideration. Concentration studies by sedimentation equilibrium and surface plasmon resonance experiments will clarify the association kinetics between each of the C3 and C3u dimers. In summary, both C3 and C3u are largely monomeric up to 2.3 mg/ml in 137 mM NaCl, in agreement with the observation of monomeric C3d in this buffer (Li *et al*., 2008). In 50 mM NaCl, greater amounts of dimer are seen, especially for C3u, in a reversible self-association with monomer (Figures 5.3-5.5). Self-association was also observed with C3d in this buffer, which is equivalent to the TED domain (Li *et al*., 2008). The interaction between factor H and C3d is seen to be strengthened in 50 mM NaCl compared to 137 mM NaCl buffer, even though this interaction is complicated by its multimeric nature (Okemefuna *et al*., 2009a). It is expected that the factor H-C3u interaction will be likewise increased in 50 mM NaCl, however an understanding of this will require dissociation constant K_D determinations for both C3u dimerisation and the factor H-C3u interaction. The larger dimerisation change seen with C3u, but not C3, is explained by the existence of a more exposed TED domain in the C3u solution structure (Figure 5.15(a)). Relatively low amounts of C3 and C3u dimer are also seen (Figure 5.5). A different second dimerisation site in C3 and C3u is suggested from the crystal structure of the C3b-SCIN complex (where SCIN is a bacterial *Staphylococcus aureus* protein that inhibits C3b) (Rooijakkers *et al*., 2009). SCIN induced the formation of a non-covalent C3b dimer through interactions between the MG4-MG5 domains that is stabilised by SCIN at this dimer interface. The modelling of our sedimentation data for the dimer based on this crystal structure account for the low amount of dimer in 137 mM NaCl. We speculate that SCIN promotes inhibitory C3b dimers starting from low amounts of dimer as part of its pathogenic inhibitory mechanism. It is possible that C3 and C3u dimerisation may offer opportunities for therapeutic interventions to reduce unwanted C3 activation.

Many C3 and C3b crystal structures were crystallised from low salt buffers (Table 5.4) using a diverse range of additives. Here, we have shown that alterations in buffer conditions affect the spatial arrangement of the 13 domains in the structures of C3 and C3u. Interestingly, examination of the crystallisation buffers for all four similar C3b crystal structures showed that comparatively low NaCl concentrations up to 50 mM

Table 5.4 C3 and C3b crystal buffers.

Protein	species	PDB code	Buffer before crystalisation	Crystallisation	Authors
C3	Human	2A73	10 mM Tris, pH 7.4	hanging drops from mother liquor containing 18% w/v PEG-3000 and 200 mM $LiNO_3$ at 20°C	Janssen *et al*., 2005
C3	Bos	2B39	20 mM Hepes (pH 7.2),100 mM NaCl	heavy-atom-containing 100 mM HEPES (pH 7.7) 40 % (w/v) PEG 400	Fredslund *et al*., 2006
C3b	Human	2I07	10 mM Tris, pH 7.4	sitting drops from mother liquor containing 9% w/v PEG-monomethylether 2000, 100 mM sodium acetate, 10 mM taurine and 50 mM Bis-Tris propane pH 7.8 at 30 °C	Janssen *et al*., 2006
C3b	Human	2ICF	25 mM Hepes, 50 mM NaCl, pH 7.2.	12% PEG 20000, 0.1 M MES, pH 6.0	Wiesmann *et al*., 2006
C3b (BbSCIN)	Human	2WIN	500 mM imidazole in 10 mM Tris, 40 mM NaCl and 5 mM $MgCl_2$, pH 8.0	75 mM sodium-potassium tartrate, 8.0% (w/v) PEG 3350 and 50 mM Bis-Tris propane, pH 6.5.	Rooijakkers *et al*., 2009
C3b	Human	2WII	10 mM Tris, 50 mM NaCl, pH 7.4	7.0% (wt/vol) PEG 3,350 and 70 mM ammonium acetate, pH 7.1	Wu *et al*., 2009
C3b	human	2HR0	20 mM Tris-HCl pH 7.5, and a protein inhibitor cocktail	200 mM Tris-HCl pH 7.5, 100 mM NaCl, 20 mM LiCl and 15% PEG 6000.	Ajees *et al*., 2006

were used, while the deviant C3b crystal structure was determined using 100 mM NaCl. Even though the deviant crystal structure is disputed (Ajees *et al*., 2006; Janssen *et al*., 2007; Ajees *et al*., 2007; Borrell, 2009), the higher NaCl concentration used for these crystals apparently resulted in a more extended TED/CUB domain arrangement that gave a better account of the C3u scattering and ultracentrifugation data in 137 mM NaCl. The resolution of separate monomer and dimer peaks by ultracentrifugation showed that the arrangement of 13 domains in both the C3 and C3u monomers is slightly more compact in 50 mM NaCl than in 137 mM NaCl. Functionally-related proteins such as Factor H with 20 SCR domains also demonstrated conformational changes with change of salt concentration (Okemefuna *et al*., 2009a). Accordingly functional studies performed in 50 mM NaCl buffer should be interpreted with caution. Buffers containing 137 mM NaCl will maintain C3 and C3u as largely monomeric proteins in physiologically-appropriate concentrations.

Conformational and self-association variability will also apply to C3 homologues, most notably C4 and C5, where this may have significant roles in complement mechanisms during their activation, and the formation of the C5 convertase by the covalent attachment of a C3b molecule to the C3bBb complex (Walport, 2001; Law & Reid, 1995). Thus the degree of self-association of C3 and C3u and conformational changes in these is important for both physiological function and laboratory investigations. The appropriate choice of buffer will be crucial to study their interactions with physiological ligands or pathogenic proteins.

(5.4) Materials and methods

(5.4.1) Purification of C3 and C3u

Wild-type C3 was purified from fresh human plasma essentially as described previously with slight modifications (Dodds, 1993). 50 ml fresh human plasma was centrifuged after adding 5% (w/v) polyethylene glycol (PEG 3500 molecular weight, Sigma Chem. Corp), and again after adding 12% PEG 3500 MWCO to the supernatant. The final pellet was resuspended in the starting buffer, formed from 95% of Buffer A (20 mM HEPES, 50 mM ε-aminocaproic acid, 5 mM EDTA, 0.5 mM Pefabloc-SC, pH 7.5) and 5% of Buffer B (as buffer A, but with the addition of 1 M NaCl). 5 ml of the filtered suspension was loaded onto a 20 ml Q-Sepharose fast flow anion-exchange column (Amersham Biosciences, GE Healthcare, Uppsala, Sweden). This column binds to proteins according to their net surface charge, and proteins are eluted according to

their isoelectric point (pI). After washing unbound proteins using 60 ml (3 column volumes) of 92% buffer A and 8% buffer B, C3 was eluted using a 120 ml (6 column volumes) linear salt gradient to 300 mM NaCl. Reduced SDS-PAGE on the C3 fractions (3.5 ml each) confirmed the α chain (115 kDa) and β chain (75 kDa). The C3 fractions from 210 mM NaCl to 250 mM NaCl were pooled to approximately 14 ml, then this was diluted 2.5 fold with Buffer A to reduce the NaCl level to 100 mM. C3 was applied onto a MonoQ 5/50 GL column (Pharmacia GE Healthcare, Uppsala, Sweden) in four to six separated batches to avoid overloading, where the column was pre-equilibrated in 90% Buffer A and 10% Buffer B. Using a salt gradient from 100 mM NaCl to 300 mM NaCl, C3 was eluted at 28 ml about 210 mM NaCl to 240 mM NaCl and checked by reducing SDS-PAGE.

For all ultracentrifugation and scattering experiments, C3 was dialysed into phosphate buffer saline (PBS: 8.1 mM Na_2HPO_4, 1.5 mM KH_2PO_4, 2.7 mM KCl, 50 mM or 137 mM NaCl, pH 7.4) overnight at 4ºC, and concentrated using Amicon® Ultra-15 centrifugal filter devices with a molecular weight cut-off of 50 kDa at a speed of 2,500 g. The C3 concentrations were determined using an absorption coefficient of 9.4 (1%, 280 nm, 1 cm path length) calculated from its composition (SWISSPROT code P01024) assuming the presence of three high mannose type oligosaccharides at Asn63, Asn917 and Asn1597 (Perkins & Sim, 1986; Hase *et al*., 1985; Hirani *et al*., 1986). The sequence-predicted molecular mass of C3 and C3u was 189,000 Da (2.3% carbohydrate), to be compared with that of 179.3 kDa for C3b. A partial specific volume of 0.739 ml/g was calculated for C3 and C3u from its composition (Perkins, 1986).The buffer density of 137 mM NaCl and 50 mM NaCl PBS were measured at 20^{o}C using an Anton-Paar DMA5000 density meter to be 1.00542 g/ml and 1.00171 g/ml, in good agreement with the predicted values of 1.00543 g/ml and 1.00175 g/ml from SEDNTERP (Laue *et al*., 1992).

Haemolytically inactive C3 (C3u, also known as C3(H_2O)) was produced by incubating C3 with 100 mM hydrazine at 37ºC for 1 hr (Sim *et al*., 1981), and then left at 4 ºC overnight. To separate C3u from any unconverted C3, C3u in 90% Buffer A and 10% Buffer B was applied to the MonoQ column, and eluted using the same procedure described for C3. C3u eluted earlier than native C3 at about 190 to 200 mM NaCl. C3u but not C3 was active in functional assays using factor I and factor H (Nan *et al*., 2008a).

(5.4.2) Analytical ultracentrifugation data for C3 and C3u

Sedimentation velocity experiments were performed at 20°C using two Beckman XL-I analytical ultracentrifuges (Beckman-Coulter Inc, Palo Alto, CA) at rotor speeds of 40,000 r.p.m. and 50,000 r.p.m.. An eight-hole AnTi50 rotor was used with standard double-sector cells with column heights of 12 mm at 20ºC, monitoring sedimentation using absorbance optics at 280 nm and interference optics. C3 was studied in a concentration series of 0.26 mg/ml to 2.28 mg/ml (PBS 50 mM NaCl) and 0.48 mg/ml to 2.20 mg/ml (PBS 137 mM NaCl). C3u was studied between 0.35 mg/ml to 2.13 mg/ml (PBS 50 mM NaCl) and 0.27 mg/ml to 2.36 mg/ml (PBS 137 mM NaCl). The continuous $c(s)$ analysis method was used to determine the $s_{20,w}$ values of C3 and C3u using SEDFIT software (version 11.7) (Dam & Schuck, 2004; Schuck, 2000). The final $c(s)$ fits were determined using a resolution of 200 and by floating the meniscus, frictional ratio, and baseline and holding the cell bottom, partial specific volume and solvent density fixed until the overall root-mean-square deviations and visual appearance of the fits were satisfactory (Figure 5.3).

(5.4.3) X-ray scattering data for C3 and C3u

X-ray solution scattering data for C3 and C3u were collected in five sessions on the camera at the ID02 high brilliance beamline at the European Synchrotron Radiation Facility (ESRF) at Grenoble, France with a ring energy of 6.0 GeV (Narayanan *et al*., 2001). In the first and second sessions, data were collected for C3 in 16-bunch mode using beam currents of 67-77 mA and 68-86 mA. In the third, fourth and fifth sessions, data were collected for C3 and C3u using beam currents of 30-37 mA (4-bunch mode), 69-83 mA (16-bunch mode) and 30-43 mA (4-bunch mode). Data were acquired using an improved fibre optically-coupled high sensitivity and dynamic range CCD detector (FReLoN) with a smaller beamstop. The sample-to-detector distance was 2.0 m. Sample flow cells were used to eliminate radiation damage in conjunction with on-line checks to confirm the absence of radiation damage by optimisation of the exposure times. Each sample was measured using four sets of ten time frames, each of length 0.05, 0.08, 0.1, 0.15, 0.2 or 0.5 sec for C3 and 0.05, 0.08, 0.1, 0.15, 0.2 or 0.25s for C3u, then these were averaged. The sample temperature was 20°C. C3 in PBS 50 mM NaCl was studied at seven concentrations between 0.18 mg/ml and 0.98 mg/ml. C3 in PBS 137 mM NaCl was studied at 18 concentrations between 0.21 mg/ml and 1.95 mg/ml. C3u in PBS 50 mM NaCl was studied at seven concentrations between 0.32 mg/ml and 1.67 mg/ml. C3u in PBS 137 mM NaCl was studied at seven concentrations between 0.28 mg/ml and 1.12 mg/ml.

In a given solute-solvent contrast, the radius of gyration R_G is a measure of structural elongation if the internal inhomogeneity of scattering densities within the protein has no effect. Guiner plot at low Q values gives the R_G and scattering at zero angle $I(0)$ (Glatter & Kratky, 1982):

$$\ln I(Q) = \ln I(0) - R_G^2 Q^2/3$$

This expression is valid in a $Q.R_G$ range up to 1.5. For elongated structure, at larger Q values than those used for the R_G analysis, the radius of gyration of the cross-sectional structure R_{XS} and the mean cross-sectional intensity at zero angle $[I(Q).Q]_{Q\to 0}$ parameters are obtained from:

$$\ln [I(Q).Q] = \ln [I(Q).Q]_{Q\to 0} - R_{XS}^2 Q^2/2$$

The R_G and R_{XS} analyses were performed using an interactive PERL script program SCTPL7 (J. T. Eaton and S. J. Perkins, unpublished software) on Silicon Graphics OCTANE Workstations. Indirect Fourier transformation of the full scattering curve $I(Q)$ in reciprocal space gives the distance distribution function $P(r)$ in real space. This yields the maximum dimension of the macromolecule L and its most commonly occurring distance vector M in real space:

$$P(r) = \frac{1}{2\pi^2} \int_o^\infty I(Q)\, Qr \sin(Qr)\, dQ$$

The transformation was implemented using GNOM software (Semenyuk & Svergun, 1982). For C3, up to 345 data points were used for Q values between 0.13 nm^{-1} and 1.73 nm^{-1}, while for C3u, up to 373 data points were used for Q values between 0.16 nm^{-1} and 1.85 nm^{-1}.

(5.4.4) Neutron scattering data for C3 and C3u

Neutron scattering data for C3 and C3u were collected in one session on the SANS2D instrument at the pulsed neutron source at ISIS at Rutherford Appleton Laboratory (RAL) in Didcot, U.K (Heenan *et al.*, 2006). Time-averaged flux is over 10^6 cm^{-2} s^{-1} (at 10Hz, 40 uA 800 MeV proton beam). Data acquisition were measured at 20°C for C3 at 2.1 mg/ml and C3u at 1.7 mg/ml.

(5.4.5) Constrained scattering and sedimentation coefficient modelling for C3 and C3u

The scattering modelling of human C3 was based on its unmodified crystal structure (PDB code 2A73) (Jansen *et al.*, 2005). The constrained scattering modelling of C3u was based on four crystal structures for human C3b (PDB codes 2I07, 2ICF,

2WIN, 2WII) (Janssen *et al*., 2006; Wiesmann *et al*., 2006; Wu *et al*., 2009; Rooijakkers *et al*., 2009). Inspection of these six crystal structures indicated that C3 and C3b could be considered as three subunits, namely the C3c region and the CUB and TED domains. These were joined by two pairs of linker peptides (1-4 in Figure 5.9(a)). Linkers 1 and 2 connect the MG7 and MG8 domains respectively in C3c with the CUB domain. Linkers 3 and 4 connect the CUB and TED domains. Their sequences are as follows: linker 1, 908-KVVPEGI-914; linker 2, 1328-AKAKDQLTC-1336; linker 3, 964-PVAQMTEDAV-973; linker 4, 1263-DAPDHQEL-1270 (C3 sequence numbering (Janssen *et al*., 2005)). Each linker was created as an extended β-strand structure, then the residues within each linker were conformationally randomised, using INSIGHT II 98.0 molecular graphics software (Accelrys, San Diego, CA, USA) on Silicon Graphics OCTANE Workstations, following Method 2 used for modelling human IgA1 (Boehm *et al*., 1999). For linker 1, 6,000 conformers were created in which the distance between the first and last α-carbon atoms ranged between 1.21 nm to 2.27 nm (Figure 5.16a). For linker 2, the 4,650 conformers ranged in length between 0.59 nm to 2.72 nm (Figure 5.16b). For linker 3, the 6,000 conformers ranged in length between 1.35 nm to 3.38 nm (Figure 5.16c). For linker 4, the 4,200 conformers ranged in length between 0.86 nm to 2.89 nm (Figure 5.16d). Five types of conformationally randomised C3u models were formed by combining the linkers and domain structures (Figure 5.9). In the first of these, the crystallographically-observed linkers 2, 3 and 4 were fixed, the connection between linker 2 and the MG8 domain was severed, and the 6,000 linker 1 conformations was used to join the TED and CUB domains to the C3c region in randomised orientations to create the C3u models. The second set of 4,650 models employed the linker 2 conformations to assemble C3u models. The third and fourth sets of models varied the position of the TED domain relative to the CUB and C3c regions by employing the 6,000 linker 3 or 4,200 linker 4 conformations respectively. A fifth set of 8,000 models was created from a randomised selection of linkers 1 and 3 conformations while holding linkers 2 and 4 fixed.

Each full C3 or C3u molecular model was used to calculate its X-ray scattering curve for comparison with the experimental C3u curve in PBS 137 mM NaCl (Table 5.3) using Debye sphere models (Perkins & Weiss, 1983). A cube side length of 0.540 nm in combination with a cutoff of 4 atoms was used to convert the atomic coordinates into Debye sphere models with 1557 spheres (C3) or 1478 spheres (C3b) that corresponded to the unhydrated structure. The hydration shell corresponding to 0.3 g H_2O/ g protein was

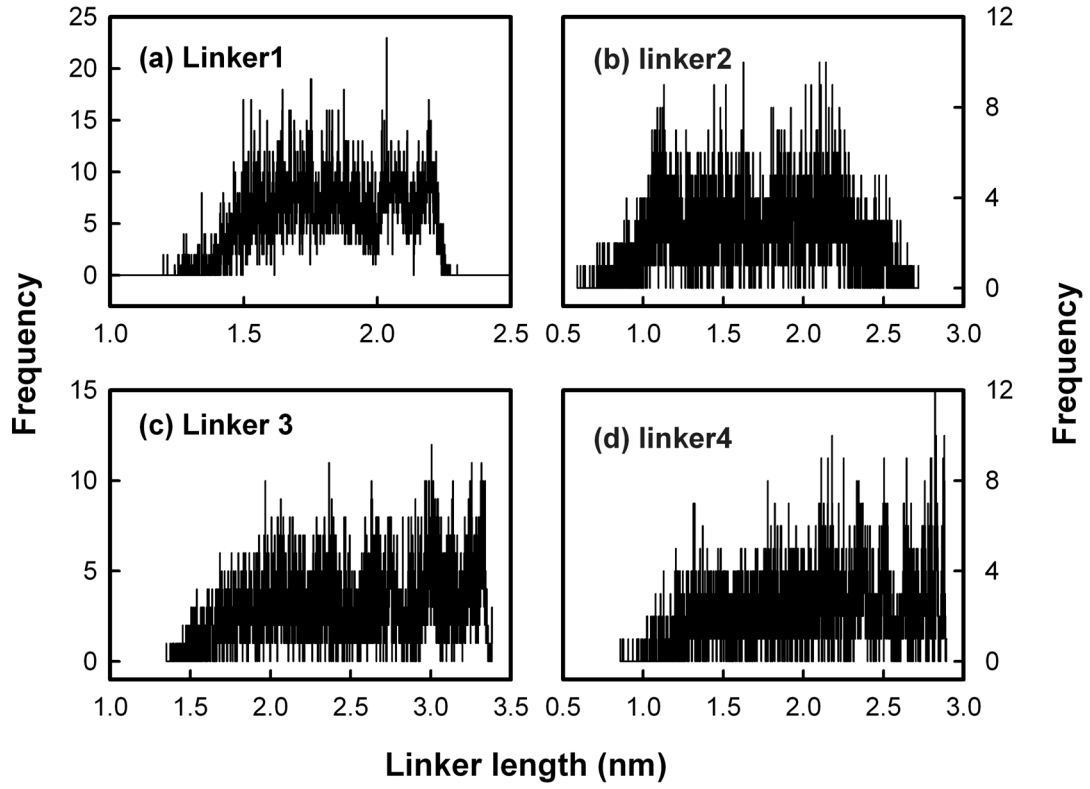


Figure 5.16 Linkers are created with the different lengths. The linkers were created and the distributions of linker lengths were checked to ensure that a sufficient range of lengths had been studied. The frequency is the number of linkers that were created for each linker length.

created using HYPRO (Ashton *et al*., 1997), where the optimal total of hydrated spheres in the C3 or C3u model is 2047 (C3) or 1943 (C3b) (Table 5.3). The X-ray scattering curve $I(Q)$ was calculated using the Debye equation adapted to spheres (Perkins & Weiss, 1983). No instrumental corrections were applied because these are considered to be negligible for the pinhole optics in use. No ad-hoc assumptions were required in relation to hydration effects, and details are given elsewhere (Perkins, 2001b). Steric overlap between the TED, CUB and C3c regions in the models was assessed using the number of spheres N in the models after grid transformation, where models showing less than 95% of the required total of 1557 or 1478 spheres were discarded. Over 54% of the 10,000 models met this "absence of steric overlap" criterion Next, the X-ray R_G and R_{XS} values were calculated from the modelled curves in the same Q ranges used for the experimental Guinier fits. This allowed for any approximations inherent in the use of the $Q.R_G$ fit range up to 1.5. Models that passed R_G and R_{XS} filters of ± 5% were then ranked using a goodness-of-fit R factor defined by analogy with protein crystallography and based on the experimental curves in the Q range extending to 1.73 nm^{-1} (C3) or 1.85 nm^{-1} (C3u) (Table 5.3; Figure 5.8).

Sedimentation coefficients $s^0_{20,w}$ were calculated directly from the hydrated Debye sphere models using the program HYDRO as a confirmation of the GENDIA analysis. (Garcia de la Torre *et al*., 1994). They were also calculated from the atomic coordinates in the HYDROPRO shell modelling program using the default value of 0.31 nm for the atomic element radius for all atoms to represent the hydration shell (Garcia de la Torre *et al*., 2000). Previous applications of these calculations to proteins are reviewed elsewhere (Perkins *et al*., 2009).

(5.5) Protein Data Bank accession number

The five best-fit C3u models for 137 mM NaCl PBS were deposited in the Protein Data Bank with the accession code 3MMQ.

Chapter Six

Solution structures reveal that complement C3u interacts independently with complement Factor H at two sites: implications for complement regulation

(6.1) Introduction

C3 is the central protein of the complement system of innate immunity (Law & Reid, 1995). In the tick-over mechanism of the complement alternative pathway, C3 is slowly hydrolysed into C3u (also known as $C3_{H2O}$) through hydrolysis of its thioester group within its thioester domain (TED). C3u shows functional similarities to C3b, the immunologically active form of C3 in which a small C3a domain is removed by cleavage. Thus C3u is able to form a C3 convertase by interactions with the Bb fragment of factor B, and this triggers the uncontrolled formation of C3b. The C3-to-C3b conversion is controlled by several regulators including factor H (FH) in order to ensure appropriate targeting against pathogens while minimising damage to host cells. FH regulates the activation of the alternative pathway by competing with factor B for binding to C3b, enhancing the dissociation of the C3bBb complex, and acting as a cofactor for the protease factor I, which leads to the irreversible proteolytic inactivation of C3b through the formation of the iC3b, C3d (the TED domain) and C3c fragments (Weiler *et al*., 1976; Pangburn *et al*., 1977).

C3 circulates in plasma at 1.0-1.5 mg/ml and FH circulates at 0.235 – 0.81 mg/ml (Law & Reid, 1995; Saunders *et al*., 2006). C3 is constructed from eight macroglobulin (MG) domains, together with the TED and C3a anaphylatoxin (ANA) domains and three other domains termed CUB, LNK and C345C (Figure 6.1(a)), while FH consists of 20 short complement regulator (SCR) domains, each containing about 60 amino acid residues (Figure 6.1(b)). Multiple functional sites exist in both C3 and FH (Alsenz *et al*., 1984, 1985; Sharma & Pangburn, 1996; Pangburn *et al*., 1991; Gordon *et al*., 1995; Ram *et al*., 1998a). Three binding sites for FH on C3b have been identified (Jokiranta *et al*., 2000). The first site is illustrated by the crystal structure of intact C3b with FH SCR-1/4, which showed that these four SCR domains bind in the crevice between the TED/CUB and the MG domains, with the SCR-1 and SCR-4 domains positioned adjacent to the C345C and MG1 domains respectively (Wu *et al*., 2009). This location is consistent with biochemical data where SCR-1/4 displayed decay-accelerating activity and this interaction is destroyed by the cleavage of C3b to C3c and C3d (Alsenz *et al*., 1984; Kühn & Zipfel, 1996). The second site involves FH binding to the C3c region (Jokiranta *et al*., 2000). This site is weaker than the other two and may correspond to SCR-6/8, and its existence has been disputed (Schmidt *et al*., 2008). The third site corresponds to the C-terminal SCR-19/20 domains which interact with the C3d fragment (TED domain) (Sharma & Pangburn, 1996; Jokiranta *et al*., 2000;

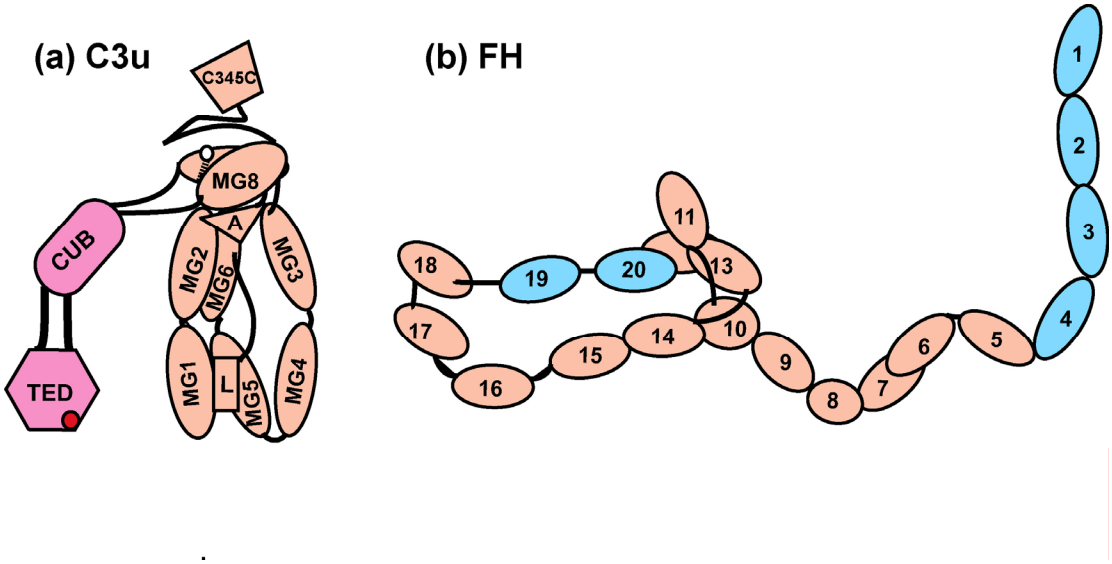


Figure 6.1 Schematic cartoon of C3u and FH. In C3u, the CUB and TED domains are shown in pink, while the remaining 11 domains (MG1-MG8, ANA (A), LNK (L) and C345C) are shown in orange (Chapter 5). In FH, the 20 SCR domains are shown in orange, with those that bind to C3b denoted in blue (SCR-1/4 and SCR-19/20).

Hellwage *et al*., 2002; Bhattacharjee *et al*., 2010). The third binding site was initially revealed by the deletion of residues 1187-1214 in C3d, which partially inhibited the FH-C3b interaction (Lambris *et al*., 1988). Residues involved in the SCR-19/20 interaction with C3d have been identified (Hellwage *et al.,* 2002; Sánchez-Corral *et al.,* 2002; Manuelian *et al.,* 2003; Jozsi *et al.,* 2006; Lehtinen *et al.,* 2009; Jokiranta *et al.,* 2006). The affinity of FH for C3b is stronger at SCR-19/20 than at SCR-1/4 (Schmidt *et al*., 2008). SCR-20 has other roles. It enables FH to adhere to glycosaminoglycan polyanionic markers on host cell surfaces (Pangburn *et al*., 1991; Blackmore *et al*., 1996, 1998; Prodinger *et al*., 1998; Ram *et al*., 1998b). These functions are reflected by the distribution of mutations, where SCR-1/4 and SCR-19/20 present most of the disease-linked sequence variations reported so far in FH (Saunders *et al*., 2007).

A molecular mechanism for the interaction between C3b and FH will clarify how the alternative pathway is regulated. In 150 mM NaCl buffer, the dissociation constant K_D is 0.59-1.6 μM for the C3b-FH interaction (Schmidt *et al*., 2008). The K_D value is 11 μM and 9.8-13.5 μM for the C3b interaction with SCR-1/4 (Wu *et al.,* 2009; Schmidt *et al*., 2008), 3.5-4.5 μM and 5.4 μM for the C3b interaction with SCR-19/20 (Schmidt *et al*., 2008; Jokiranta *et al*., 2000, 2006), and 29 μM for the SCR-16/20 interaction with C3d (Okemefuna *et al.,* 2009b). Binding between C3b and FH is dependent on ionic strength and pH (Soames & Sim, 1997). These micromolar K_D values in physiological salt indicate that the C3b-FH interaction is comparatively weak. Only 63% complex formation will be achieved, even with the strongest K_D value of 0.59 μM in the presence of the highest reported plasma concentrations of 1.5 mg/ml (C3b) and 0.81 mg/ml (FH) (8.4 μM and 5.3 μM respectively). This is a characteristic feature of the interactions between the complement proteins (Perkins *et al.,* 2010a, 2010b), and makes structural studies of the C3b-FH complex more difficult to achieve. Recently we have described the solution structures for C3u and FH by analytical ultracentrifugation (AUC) and constrained small-angle X-ray scattering (SAXS) modelling, and identified dimeric and higher oligomeric forms of C3u and FH (Li *et al.,* 2010; Chapter 5; Nan *et al.,* 2010). These two studies enable the C3u-FH interaction to be studied. Even though full complex formation was not achieved, the joint use of AUC and SAXS unambiguously identified a 1:1 complex between C3u and FH, and we propose a solution structure for this complex in which there are no major conformational changes in FH. In addition, a higher oligomer of the C3u-FH complex with a 2:1 stoichiometry was identified, showing that C3u binds independently to two different sites on FH. In

combination with the multimeric C3d-FH complexes identified previously (Okemefuna *et al.,* 2009b), these results provide novel insights into a molecular mechanism for the C3u-FH interaction as well as that for the C3b-FH interaction.

(6.2) Results and Discussion

Following our studies of the interactions between recombinant complement receptor type 2 SCR-1/2 and SCR-1/15 with recombinant C3d, and that between plasma FH and recombinant C3d (Gilbert *et al.,* 2005; Li *et al.,* 2008; Chapter 4; Okemefuna *et al.,* 2009b), size-exclusion chromatography was first performed in order to characterise the C3u-FH complexes formed in 137 mM and 50 mM NaCl phosphate buffers. Next, AUC and SAXS were performed in order to define these mixtures quantitatively in terms of free and complexed proteins. Finally, constrained sedimentation and scattering modelling was performed in order to define the structures and K_D values for the C3u-FH complexes (Li *et al.,* 2010; Chapter 5; Nan *et al.,* 2010).

(6.2.1) Size exclusion chromatography of C3-FH and C3u-FH mixtures

Native FH, C3 and C3u were purified from plasma (Materials & Methods). C3, C3u and FH were checked by SDS-PAGE. While all proteins ran as single bands, FH at 130 kDa migrated faster than C3 or C3u at 180 kDa on a non-reducing gel (Figure 6.2(c)). The observed mass of 130 kDa is less than the expected mass of 154 kDa, and this difference is attributed to the high carbohydrate content of FH. For both the C3-FH and C3u-FH mixtures, only two bands from C3 or C3u and FH were observed in the non-reducing gel. Because no third band was observed at a mass of about 300 kDa, the interaction between C3u-FH was inferred to be weak. On a reducing gel, FH at 170 kDa migrated slower than the α-chain and β-chain of C3 and C3u at 110 kDa and 70 kDa. Weak bands for free FH at 120 kDa and 40 kDa indicated trace amounts of a previously-characterized proteolytic cleavage of FH within the disulphide-bridged region of SCR-5 (Perkins *et al.,* 1991).

When the three proteins and the C3-FH and C3u-FH mixtures were analysed by size-exclusion chromatography, weak complex formation was observed for C3u-FH but not for C3-FH. Accordingly the C3-FH mixtures were used as controls for complex formation observed with the C3u-FH mixtures:

(i) In 137 mM NaCl (Figure 6.2(a)), C3 (red dashed line) and C3u (red solid line) each eluted as single homogeneous peaks at 77 ml (Li *et al.,* 2010; Chapter 5). FH (blue line)

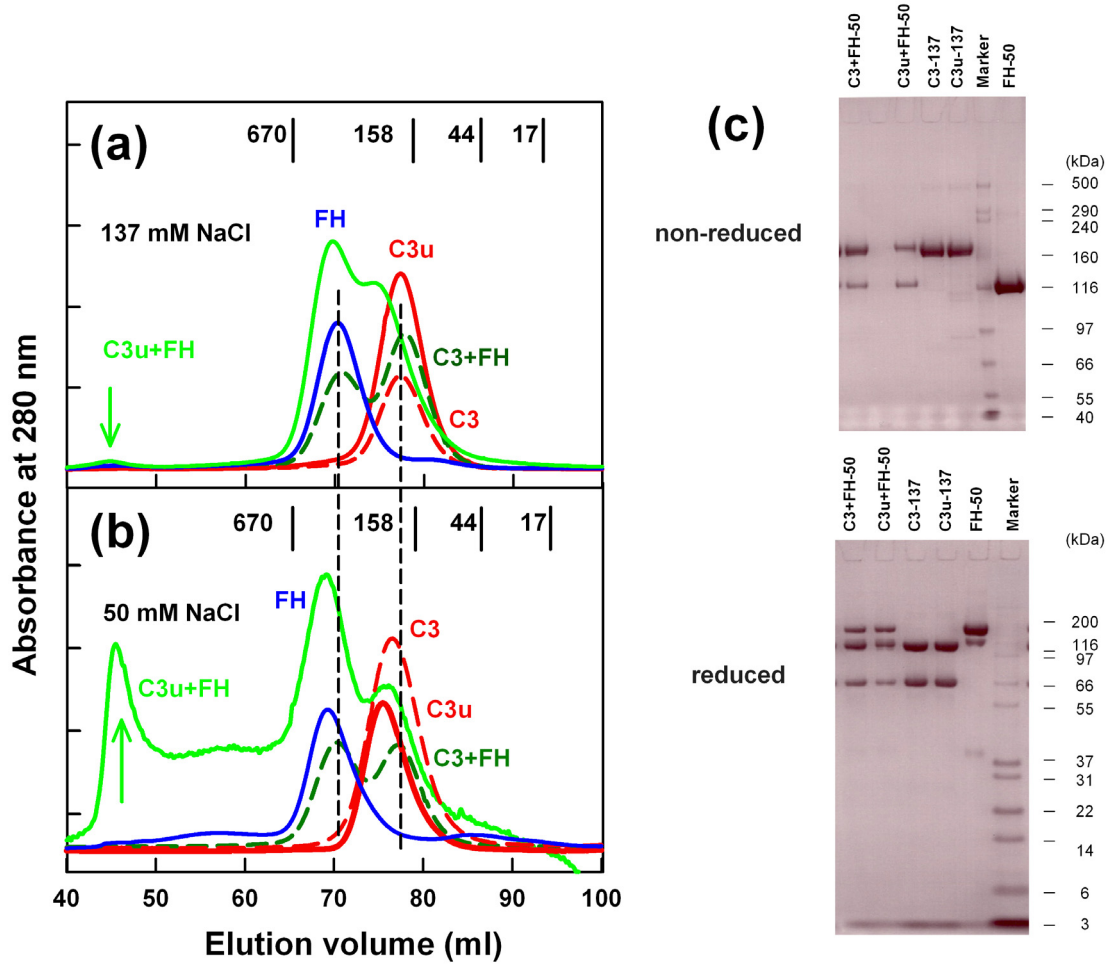


Figure 6.2 Size-exclusion chromatography and SDS-PAGE analyses of C3, C3u, FH and their mixtures. In (a,b) the elution positions of four molecular mass standards are shown in kDa, and the elution positions of C3 and FH are indicated by vertical dashed lines.

(a) In 137 mM NaCl buffer, FH (blue), C3 (red dashed) and C3u (red) are each eluted as single peaks. The C3u-FH mixture eluted as three peaks that correspond to the C3u-FH complex, free FH and partially-complexed C3u (green), while the C3-FH mixture eluted as two peaks that correspond to free FH and free C3 (green dashed).

(b) In 50 mM NaCl buffer, FH (blue), C3 (red dashed) and C3u (red) are each eluted as single peaks. The C3u-FH mixture eluted as three peaks (green), while the C3-FH mixture eluted as two peaks (green dashed).

(c) Non-reducing and reducing SDS-PAGE analyses of C3, C3u, FH and their mixtures in 50 mM NaCl and 137 mM NaCl as labelled. A precast 3-8% Tris-acetate gradient gel was used for the non-reduced analysis and a 4-12% Bis-Tris gradient gel was used for the reduced analysis (Invitrogen). The molecular mass standards are indicated in kDa at the right hand side.

eluted as a single peak at 70 ml, followed by a minor peak assigned to residual traces of human serum albumin (Nan *et al.,* 2008). The equimolar C3-FH mixture eluted as two peaks in 137 mM NaCl at the positions of unbound C3 and FH (green dashed line), thus no complex formation was observed. The equimolar C3u-FH mixture eluted as three peaks in 137 mM NaCl (green solid line). That at 45 ml corresponded to small amounts of the C3u-FH complex, which was followed by a second peak at 70 ml (FH) and a third peak at 75 ml (C3u) slightly overlap. Because the FH and C3u peaks eluted earlier than those for either proteins on their own (red and blue), this indicated that some complex formation had occurred. The excess of uncomplexed C3u and FH compared to the complex indicated that the complex is weakly formed in physiological 137 mM NaCl with a μM K_D value.

(ii) In 50 mM NaCl (Figure 6.2(b)), C3 (red dashed line) and C3u (red solid line) each eluted earlier as a single homogeneous peak at about 75 ml, and this was previously explained in terms of greater self-association of both proteins in low salt (Li *et al.,* 2010). FH likewise self-associates more in 50 mM NaCl (Okemefuna *et al.,* 2009a). This self-association was reflected by a new broad FH peak at 50-60 ml, and a FH monomer peak that eluted earlier at 68 ml (Figure 6.2(b)). The equimolar C3-FH mixture again eluted as two peaks in 50 mM NaCl at the positions of unbound C3 and FH (green dashed line), thus no complex formation was observed. The equimolar C3u-FH mixture eluted as three major peaks (green solid line). The first peak at 45 ml corresponds to increased complex formed by C3u and FH, while the second and third peaks at 68 ml and 75 ml correspond to unbound FH and C3u. The observation of three peaks again indicated a μM K_D value, and the increased amount of the C3u-FH complex in 50 mM NaCl than in 137 mM NaCl indicates a strong ionic strength dependence of complex formation, similar to that reported for the C3b-FH interaction (Soames & Sim, 1997).

(6.2.2) Sedimentation velocity data for C3-FH and C3u-FH mixtures

AUC using sedimentation velocity experiments reports on macromolecular structures in solution by analysis of the sedimentation time course (Cole *et al*., 2008). The size distribution analyses *c(s)* calculated using SEDFIT reports on the species that are present from the number of peaks observed in the *c(s)* plot. Macromolecular elongation is monitored through the sedimentation coefficient $s_{20,w}$ values. Freshly-prepared C3 was used to avoid its hydrolysis to C3u during storage. Prior to ultracentrifugation, the C3-FH and C3u-FH mixtures were subjected to gel filtration on

Superose 6 column to remove pre-existing aggregates (Figure 6.2). The experiments were done in 137 mM NaCl and 50 mM NaCl to follow the conditions used for gel filtration. Analyses of up to 300 scans in SEDFIT revealed excellent boundary fits (Figure 6.3(a,c,e,g)).

(i) In 137 mM NaCl for the C3-FH mixture, the interference and absorbance data sets gave an $s_{20,w}$ value of 5.91 S for FH, and an $s_{20,w}$ value of 8.21 S for C3 (Table 6.1; peaks 1 and 2 in Figure 6.3(b)). Additional peaks 3 and 4 were observed in Figure 6.3(b) that was not expected. This is attributed to experimental error, when the C3-FH mixture was stored at 4°C fridge eight days before the AUC run, and C3 became hydrolysed to form C3u and other degradation products that were revealed in SDS-PAGE (not shown). Thus peak 2 was in fact a mixture of peaks for both C3 and C3u, although this was not clearly resolved. Peak 3 of the C3-FH mixture was attributed to 1:1 C3u-FH complex formation, and the small peak 4 was attributed to higher 2:1 complex (see below). This experiment needs to be repeated using freshly-purified C3.

(ii) In 137 mM NaCl for the C3u-FH mixture, peaks 1 and 2 for FH and C3u were observed with $s_{20,w}$ values of 5.77 S, and 7.63 S respectively (Figure 6.3(d); Table 6.1). Peak 3 with a $s_{20,w}$ value of 9.62 S corresponded to 1:1 complex formation, while very low amounts of peak 4 corresponding to 2:1 complex formation was also observed. At a total concentration of 1.02 mg/ml, integration showed that peak 1 corresponded to 44%, peak 2 to 6% and peak 3 to 50% of the mixture. By converting the *c(s)* peak integrations to molarities, the three peaks in Figure 6.3(d) gave a K_D value of 0.59 ± 0.11 µM. This determination is similar to the previous K_D value of 0.59 ± 0.04 µM reported by surface plasmon resonance (Schmidt *et al.,* 2008). FH forms oligomers in which the $s_{20,w}$ values of the well-resolved dimer and trimer peaks are approximately at 7.6 S and 9.4 S (Nan *et al*., 2008, 2010; Okemefuna *et al*., 2009a). Because the FH dimer and trimer peaks overlap with the $s_{20,w}$ value for C3u at 7.66-7.44 S (Li *et al.,* 2010) and the $s_{20,w}$ value for the 1:1 complex at 9.62-9.65 S (Table 6.1), it was not possible to deduce whether FH dimers and trimers were removed by the C3u-FH interaction, although this is likely.

(iii) In 50 mM NaCl, the C3-FH mixture showed two peaks at 5.80 S and 7.97 S for FH and C3 respectively, in which FH accounted for 47 % and C3 account for 53 % of the mixture at 1.36 mg/ml (Figure 6.3(f); Table 6.1). The absence of other peaks indicated that no complexes were formed, in agreement with gel filtration (Figure 6.2(b); green dashed line). The baseline between 9-15 S in the *c(s)* plot was not flat, and

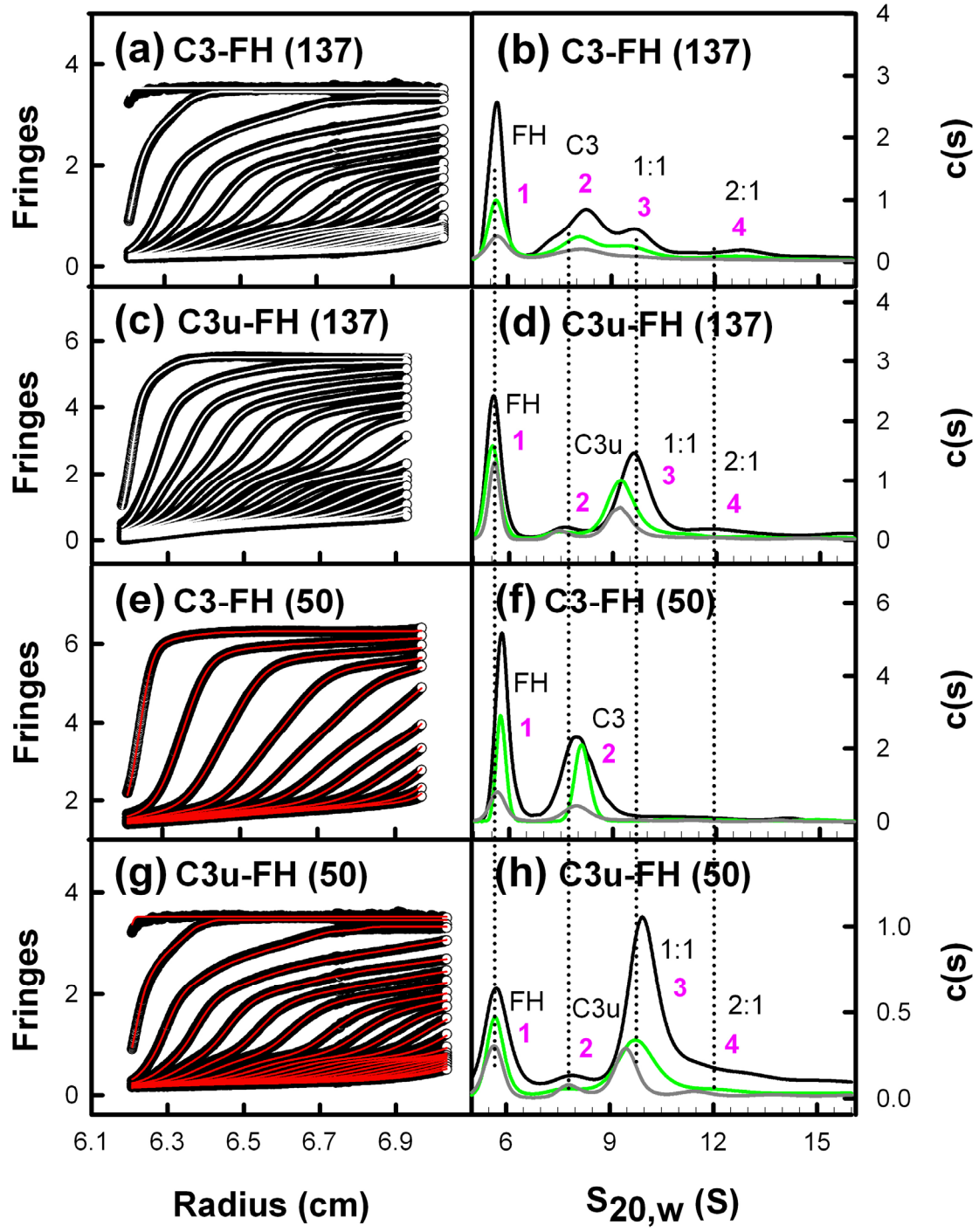


Figure 6.3 Size distribution analyses *c(s)* for mixtures of C3 and C3u with FH (Legend overleaf).

Figure 6.3 Size distribution analyses *c(s)* for mixtures of C3 and C3u with FH (continued).

(a,c,e,g) Sedimentation velocity boundary fits for C3-FH and C3u-FH mixtures at approximately 1.5 mg/ml using interference optics in 137 mM (white) and 50 mM (red) NaCl buffer. Only every third or fifth scans of the fitted boundaries are shown for reason of clarity.

(b,f) The *c(s)* distributions for C3-FH mixtures in 137 mM and 50 mM NaCl buffer. The three colours represent different concentrations. The total C3-FH concentrations range from 0.36 mg/ml to 1.07 mg/ml in 137 mM NaCl, and 0.45 mg/ml to 1.36 mg/ml in 50 mM NaCl buffer. Here and below, the vertical dotted lines show the approximate positions of the peaks for FH (denoted by 1), C3 or C3u (2), the 1:1 complex (3) and the 2:1 complex (4).

(d,h) The *c(s)* distributions for C3u-FH mixtures in 137 mM and 50 mM NaCl buffer. Individual peaks are labelled. In 137 mM NaCl, the total C3u-FH concentrations range from 0.51 mg/ml to 1.53 mg/ml, while in 50 mM NaCl, the total C3u-FH concentrations range from 0.54 mg/ml to 1.62 mg/ml.

may reflect the presence of unbound FH trimer and higher FH oligomers as well as unbound C3 dimer, although these peaks could not be resolved.

(iv) In 50 mM NaCl, the C3u-FH mixture showed four peaks 1-4 (Figure 6.3(h); Table 6.1). Peaks 1 and 2 show similar $s_{20,w}$ values to those of FH (5.88 S) and C3u (7.90 S). For the complexes, the mean $s_{20,w}$ value of peak 3 is 9.65 S, and peak 4 is 12.47 S, both of which have larger peak areas than those in 137 mM NaCl, which indicated greater complex formation (Figure 6.3(d)). A concentration dependence was seen, where the relative sizes of peaks 1 and 3 at 0.54 mg/ml was reversed at 1.62 mg/ml. Integration showed that peak 1 corresponded to 35%, peak 2 to 10%, peak 3 to 48% and peak 4 to 7% of the mixture at 0.54 mg/ml (total concentration). At 1.62 mg/ml, peak 1 corresponded to 26%, peak 2 to 8%, peak 3 to 51% and peak 4 to 15%. The change in percentages showed that more complex formation occurred at higher concentration. By converting the *c(s)* peak integrations to molarities, peaks 1, 2 and 3 in Figure 6.3(h) resulted in an approximate K_D value of 0.11 ± 0.06 μM. This decreased K_D value showed that C3u binding to FH was stronger when [NaCl] is increased.

The C3u-FH interaction in 137 mM NaCl and 50 mM NaCl were analysed using linear regression fits of the $s_{20,w}$ values for FH, C3u and the 1:1 C3u-FH complex (peaks 1, 2 and 3 in Figure 6.3).

(i) For 137 mM NaCl, the $s_{20,w}$ values of FH were unchanged from 0.18 mg FH/ml to 0.68 mg FH/ml, and give a $s^0_{20,w}$ value of 5.78 ± 0.09 S (Figure 6.4(a); Table 6.1). The molecular weight and f/f_o values for FH were 152 ± 13 kDa and 1.66 ± 0.11 respectively. For 50 mM NaCl, the $s_{20,w}$ values of FH showed an almost negligible change with concentration from 0.24 mg/ml to 0.72 mg/ml of 5.90 ± 0.25 S to 5.93 ± 0.06 S. The molecular weight and f/f_o values for FH by itself were 153 ± 16 kDa and 1.68 ± 0.25 respectively. The slightly increased $s_{20,w}$ values compared to 137 mM NaCl correspond to a NaCl-dependent conformational change. This $s_{20,w}$ change for unbound FH in the presence of C3u is similar to that seen for FH on its own when the NaCl concentration is increased (Okemefuna *et al*., 2009a).

(ii) For 137 mM NaCl, the $s_{20,w}$ values of C3u were unchanged from 0.28 mg C3u/ml to 0.85 mg C3u/ml, and give a $s^0_{20,w}$ value of 7.63 ± 0.20 S (Figure 6.4(b); Table 6.1). For 50 mM NaCl, the $s_{20,w}$ values of C3u increase from 0.30 mg/ml to 0.90 mg/ml from 8.04 ± 0.09 S to 8.28 ± 0.10 S. It is concluded that the properties of unbound C3u in the presence of FH again behaves in the same way as that reported for C3u on its own (Li *et al*., 2010; Chapter 5).

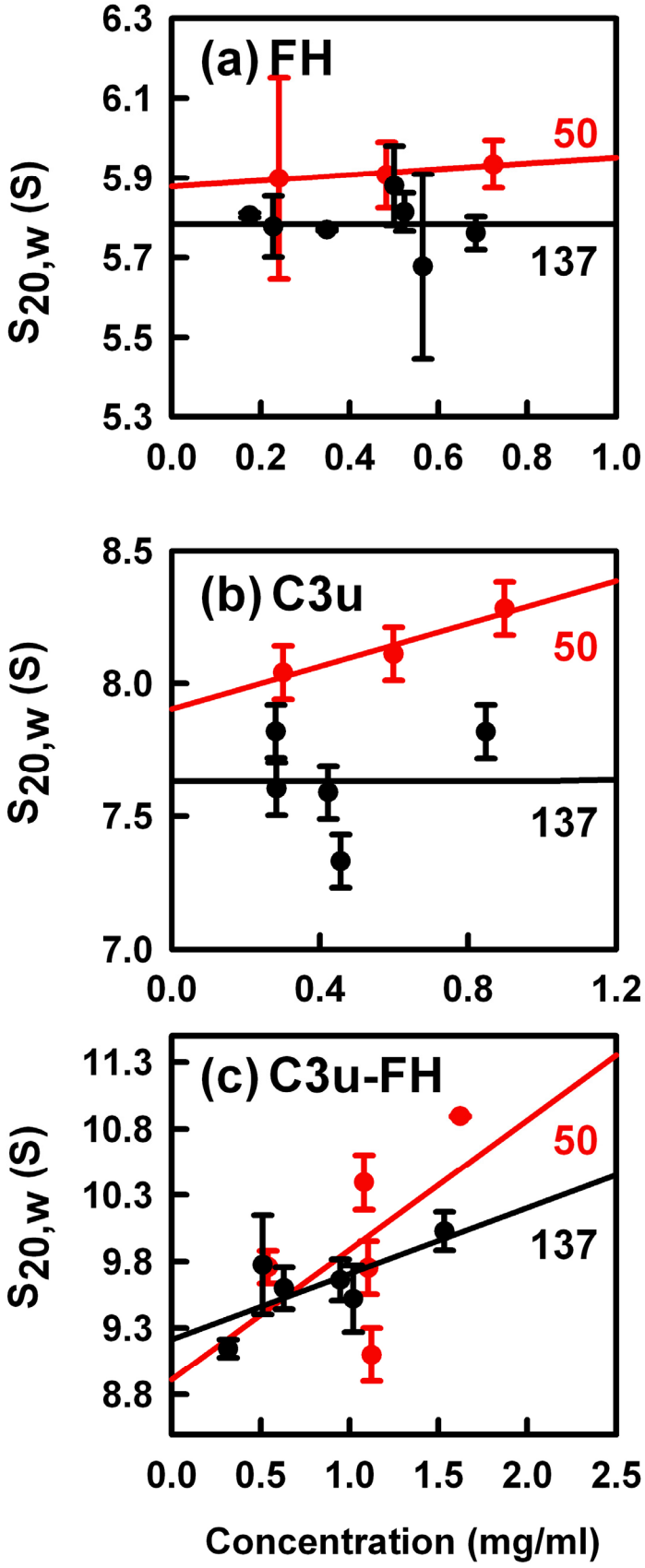


Figure 6.4 Concentration dependence of the $s_{20,w}$ values for C3u, FH and their 1:1 complex (Legend overleaf).

Figure 6.4. Concentration dependence of the $s_{20,w}$ values for C3u, FH and their 1:1 complex (continue).

(a) The $s_{20,w}$ values for FH (peak 1 in Figure 6.3(d,h)) were fitted by linear regression to result in $s^o_{20,w}$ values at zero concentration of 5.78 ± 0.09 S in 137 mM NaCl (black) and 5.88 ± 0.07 S in 50 mM NaCl (red). Here and in other places, statistical error bars are shown where visible.

(b) The $s_{20,w}$ values for C3u (peak 2 in Figure 6.3(d,h)) were fitted to result in $s^o_{20,w}$ values of 7.63 ± 0.20 S in 137 mM NaCl and 7.90 ± 0.40 S in 50 mM NaCl.

(c) The $s_{20,w}$ values for the 1:1 complex (peak 3 in Figure 6.3(d,h)) were fitted to result in $s^o_{20,w}$ values of 9.21 ± 0.50 S in 137 mM NaCl and 8.91 ± 0.98 S in 50 mM NaCl.

(iii) For the C3u-FH complex in 137 mM NaCl buffer (peak 3 in Figure 6.4(c)), even though the $s_{20,w}$ values appear to increase from 0.32 mg/ml to 1.53 mg/ml (total concentration), this increase was not greater than statistical error, although it is visible in Figure 6.3(d). For the C3u-FH complex in 50 mM NaCl, the concentration dependence of the $s_{20,w}$ values from 9.10 ± 0.20 S to 10.90 ± 0.01 S was greater. The molecular mass and f/f_o values of peak 3 in 137 mM NaCl were 333 ± 16 kDa and 1.79 ± 0.11 respectively, and those in 50 mM NaCl were 356 ± 46 kDa and 1.68 ± 0.37 respectively. If C3u and FH form a 1:1 complex, its predicted mass would be 343 kDa. The good agreement between the observed and predicted masses showed that a 1:1 complex was formed in both buffers, and that a reaction boundary between C3u and FH was not present. [**Footnote:** In *c(s)* plots, reaction boundaries are concentration-dependent peaks that correspond to the co-sedimentation of two different species if they exchange between themselves during sedimentation. They exhibit well-defined $s_{20,w}$ values and masses between the $s_{20,w}$ values and masses of the two different species, and are distinct from undisturbed boundaries which correspond to the actual sedimentation species themselves.] The concentration dependence of peak 3 in 50 mM NaCl is most easily explained by the formation of a 1:1 C3u-FH complex from C3u and FH, to which there is a small contribution from the concurrent formation of an additional 2:1 $C3u_2$-FH complex (peak 4 below). A contribution from the concentration-dependent C3u dimer formation previously observed in 50 mM NaCl may also be present (Li *et al.*, 2010).

Trace amounts of peak 4 were visible for the C3u-FH mixtures in 137 mM NaCl that were significantly increased in 50 mM NaCl (Figures 6.3(d,h)). Larger but ill-defined changes in the $s_{20,w}$ values were detectable in 50 mM NaCl. The mass of peak 4 in 137 mM NaCl was 476 ± 28 kDa, which is in good agreement with the predicted mass of 532 kDa for a $C3u_2$-FH complex. The observed mass for peak 4 in 50 mM NaCl was lower at 391 ± 47 kDa. While peak 4 can be identified as a 2:1 complex, larger amounts of the 2:1 complex is formed in 50 mM NaCl buffer, and there is a possibility that peak 4 may correspond to a reaction boundary between the 1:1 C3u-FH complex and the formation of one or more 2:1 C3u-FH complexes.

(6.2.3) X-ray scattering data for C3-FH and C3u-FH mixtures

Solution scattering is a diffraction technique that studies the overall structure of biological macromolecules in solution (Perkins *et al.*, 2008). Freshly prepared C3-FH

and C3u-FH mixtures were studied by X-ray scattering in four beam sessions (Figure 6.5). The interaction between C3 or C3u with FH was assessed by Guinier analyses of the X-ray *I(Q)* data at low Q values to give the radius of gyration R_G and forward scattered intensities *I(0)*, where $Q = 4\pi \sin\theta / \lambda$ (2θ = scattering angle; λ = wavelength). In concentration series, the Guinier fits were linear as required, in an appropriate $Q.R_G$ range up to 1.5 (Figure 6.5).

The X-ray Guinier results for the equimolar 1:1 C3-FH and C3u-FH mixtures in 137 mM NaCl (black) and 50 mM NaCl (red) buffers are summarised in Figure 6.6.

(i) In 137 mM NaCl, the R_G values for the C3-FH mixture were unchanged with protein concentration at 7.83 ± 0.58 nm (16 values) (Figure 6.6(a)). The *I(0)/c* value was unchanged at 0.0306 ± 0.0006 (Figure 6.6(b)); this is proportional to the mean molecular mass. If no complex formation occurs, the weighted mean of the R_G values of 4.52 nm for free C3 (Li *et al*., 2010; Chapter 5) and 8.90 nm for free FH (Okemefuna *et al.,* 2009a) is calculated to be 6.62 nm for an equimolar C3-FH mixture (Perkins *et al.,* 1985). Since this mean is less than the observed value, and because C3 and FH do not interact with each other (Figure 6.3(f)), this R_G difference is inferred to reflect the propensity of each of C3 and FH to form trace oligomers with increased R_G values, rather than the formation of complexes.

(ii) In 137 mM NaCl, the R_G values for the C3u-FH mixtures were also unchanged with concentration at 8.83 ± 0.34 nm (16 values) (Figure 6.6(c)), but are now larger than that of 7.83 nm for C3-FH. The *I(0)/c* values were also larger at 0.0460 ± 0.0023 for C3u-FH (Figure 6.6(d)). Because the R_G value of 8.83 nm is noticeably larger than that of 6.77 nm (Perkins *et al.,* 1985) for the weighted mean of the R_G values of 4.88 nm for free C3u and 8.90 nm for free FH, this increase reflected the formation of C3u-FH complexes.

(iii) In 50 mM NaCl, the R_G values of the C3-FH mixture in 50 mM NaCl increased linearly (Figure 6.6(a)). At zero concentration, the R_G values were extrapolated to 8.07 ± 1.22 nm, which is the same as that in 137 mM NaCl. The *I(0)/c* values of the C3-FH mixture also increased linearly. At zero concentration, the *I(0)/c* values extrapolated to 0.0400 ± 0.0090, which was slightly larger than that in 137 mM NaCl. Overall, the changes were attributed to increased oligomer formation for C3 seen previously in 50 mM NaCl (Li *et al*., 2010; Chapter 5). No complex formation between C3 and FH was deduced, in agreement with Figure 6.3(f).

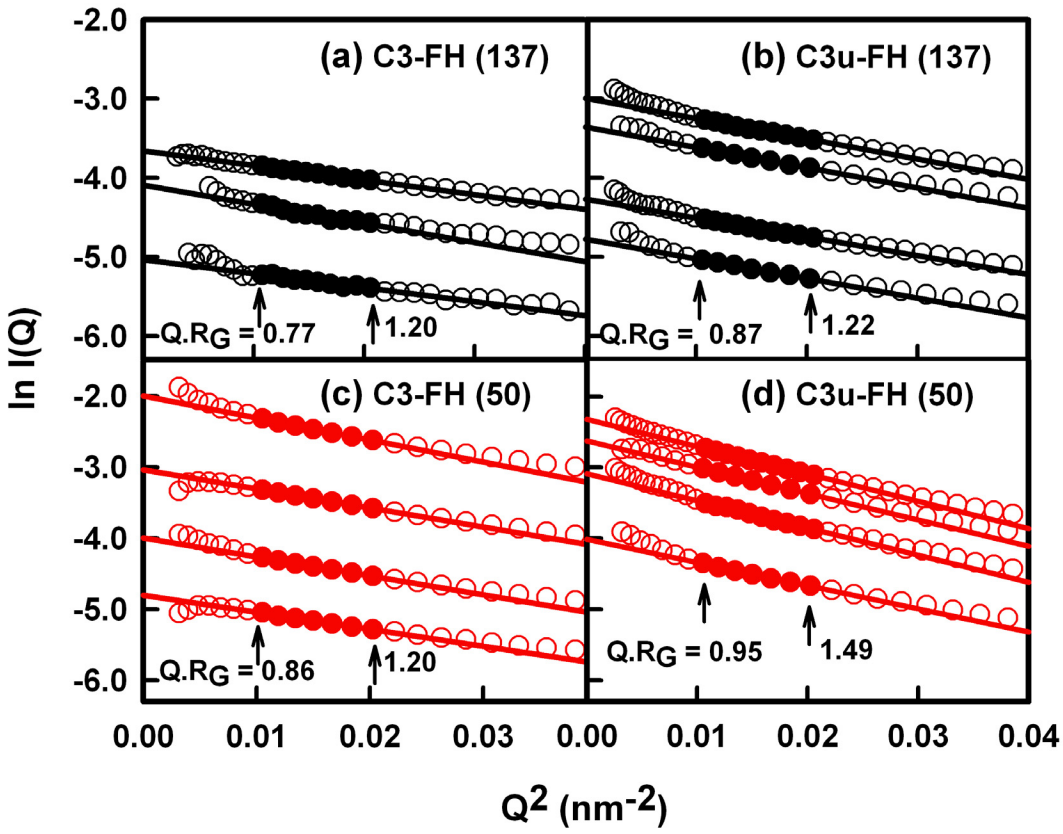


Figure 6.5 X-ray Guinier R_G analyses of C3-FH and C3u-FH mixtures. The Q range used for the R_G fits was 0.10 to 0.14 nm^{-1}. The data points are denoted by open circles; the filled circles were used to determine the R_G values based on the best fit lines as shown.

(a,c) Plots for 1:1 mixtures of C3-FH are shown at total concentrations between 0.30 mg/ml to 0.78 mg/ml in 137 mM NaCl (black) and 0.33 mg/ml to 1.32 mg/ml in 50 mM NaCl (red) from bottom to top.

(b,d) Plots for 1:1 mixture of C3u-FH are shown at total concentrations between 0.35 mg/ml to 1.04 mg/ml in 137 mM NaCl and 0.47 mg/ml to 1.04 mg/ml in 50 mM NaCl.

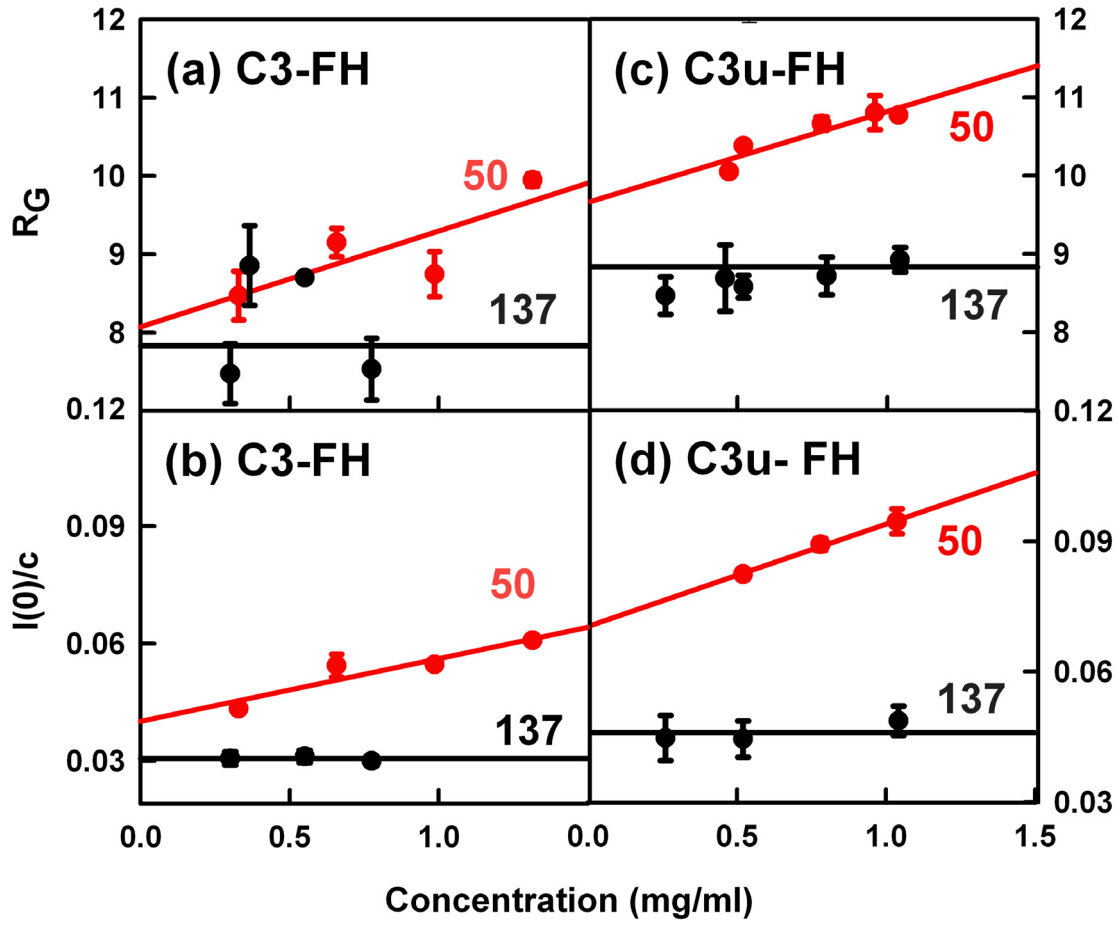


Figure 6.6 Concentration dependence of the Guinier R_G and $I(0)/c$ values for C3-FH and C3u-FH mixtures. Each value was measured in quadruplicate and averaged for linear regression fits. The fitted lines shown for 137 mM NaCl buffer are the mean value. Statistical error bars are shown where visible.

(a,c) The R_G values are shown for 137 mM NaCl (black) and 50 mM NaCl buffer (red). For C3-FH in 50 mM NaCl, measurements of 8.47 ± 0.31 nm at 0.33 mg/ml (4 values), 9.15 ± 0.18 nm at 0.66 mg/ml (4 values), 8.75 ± 0.29 nm at 0.99 mg/ml (4 values), and 9.95 ± 0.08 nm at 1.31 mg/ml (4 values) are shown. For C3u-FH in 50 mM NaCl, measurements of 10.05 ± 0.06 nm at 0.47 mg/ml (4 values), 10.38 ± 0.02 nm at 0.52 mg/ml (4 values), 10.67 ± 0.09 nm at 0.78 mg/ml (4 values), 10.80 ± 0.22 nm at 0.96 mg/ml (4 values) and 10.78 ± 0.04 nm at 1.04 mg/ml (4 values) are shown.

(b,d) The corresponding *I(0)/c* values are shown for 137 mM and 50 mM NaCl buffer.

(iv) In 50 mM NaCl, the R_G values for the C3u-FH mixture increased linearly with concentration. The R_G value was 9.66 ± 1.16 nm when extrapolated to zero concentration. Because this R_G value is larger than that of 8.83 nm for the C3u-FH mixture in 137 mM NaCl, this shows that an increased proportion of the C3u-FH complex is formed in reduced salt conditions. The *I(0)/c* values of the C3u-FH mixture also increased linearly, which agrees with increased R_G values. The *I(0)/c* value of the C3u-FH mixture in 50 mM NaCl was 0.0702 ± 0.0235 when extrapolated to zero concentration. This value was larger than the value of 0.0460 above, and showed that additional complexes were formed in low salt buffer.

Additional X-ray data were collected for C3-FH and C3u-FH mixtures in 137 mM and 50 mM NaCl at three different molar ratios of 1:0.3, 1:0.5 and 1:1. The C3 and C3u concentration was maintained at 0.50 mg/ml. While linear Guinier plots were obtained (Figure 6.7), no further R_G or *I(0)/c* data analysis was performed for reason of its complexity. These extra data sets will be useful as controls for the constrained analyses of the C3-FH and C3u-FH mixtures in order to refine the K_D determination from scattering curve fits (see below).

The distance distribution function *P(r)* in real space leads to the determination of overall lengths *L,* following an assumption of the value of the maximum dimension D_{max}. The *P*(*r*) function also gives an independent calculation of the R_G and *I(0)* values that is based on the full *Q* range for comparison with the Guinier analyses based on low *Q* values. The four pairs of R_G values from the *P*(*r*) and Guinier analyses were in good agreement, indicating the self-consistency of the pairs of data sets (Table 6.1).

(i) In 137 mM NaCl, the *L* value for the C3-FH mixture was 28 nm (Figure 6.8(a)). The most common distance *M* for the C3-FH mixture was 5.5 nm at 0.30 mg/ml, increasing to 6.2 nm at 0.78 mg/ml (Figure 6.8(a)). Previously, the *L* values of unbound C3 and FH were reported to be 16 nm and 28-32 nm respectively in 137 mM NaCl, and the most common distance *M* for unbound C3 and FH were 5.0 nm and 5.4 nm (Li *et al*., 2010; Chapter 5; Nan *et al.,* 2008; Okemefuna *et al.,* 2009a). The lack of change in *L* for the C3-FH mixture agrees with the absence of complex formation, while the change in *M* is not considered significant. The *P*(*r*) curves thus agreed with the lack of interaction in a freshly-prepared mixture of C3-FH in 137 mM NaCl.

(ii) In 137 mM NaCl, the C3u-FH mixture showed *L* values of 30 nm and *M* values of 5.5 to 6.4 nm that increased with decrease in concentration (Figure 6.8(b)).

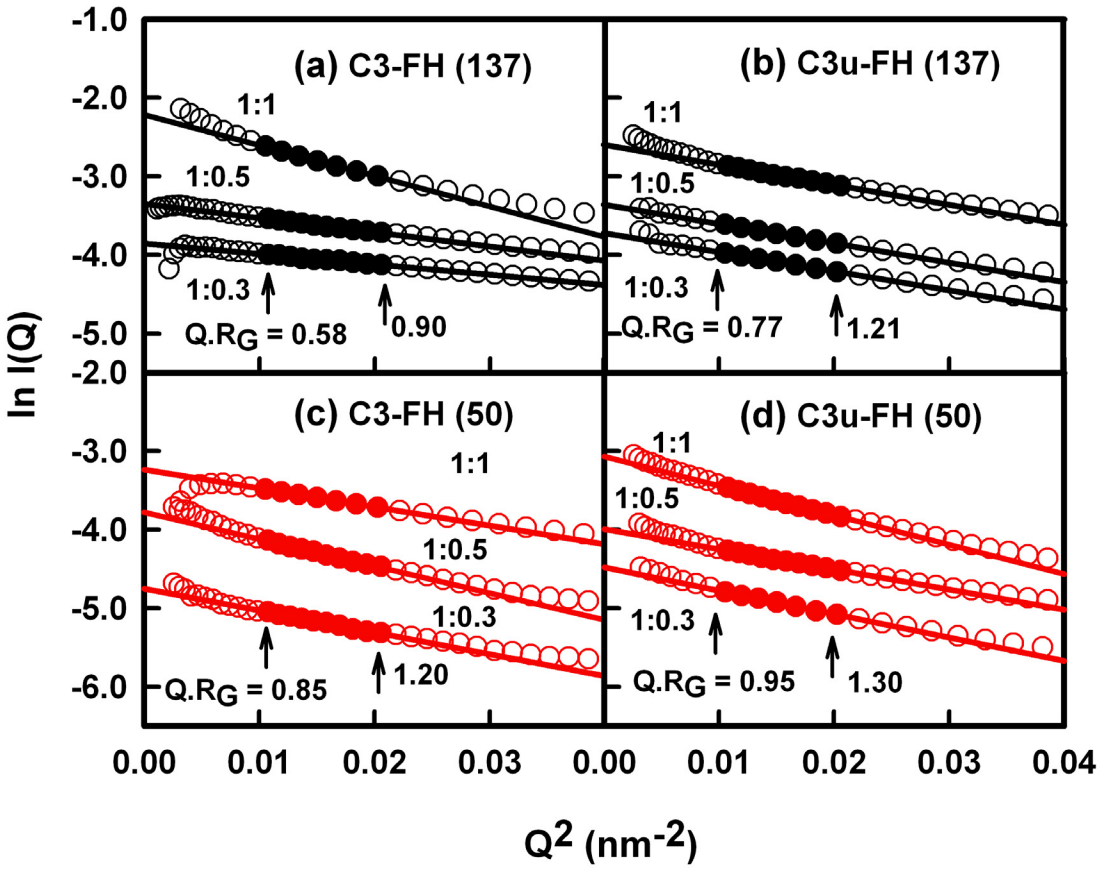


Figure 6.7 X-ray Guinier R_G analyses of C3-FH and C3u-FH mixtures in different molar ratios. The Q range used for the R_G fits was 0.10 to 0.14 nm^{-1}. The data points are denoted by open circles; the filled circles were used to determine the R_G values based on the best fit lines as shown.

(a,c) C3-FH mixtures in molar ratios of 1:0.3, 1:0.5 and 1:1 with C3 fixed at 0.5 mg/ml in 137 mM NaCl (black) and in 50 mM NaCl (red) from bottom to top.

(b,d) C3u-FH mixtures in molar ratios of 1:0.3, 1:0.5 and 1:1 with C3u fixed at 0.5mg/ml in 137 mM NaCl (black) and in 50 mM NaCl (red) from bottom to top.

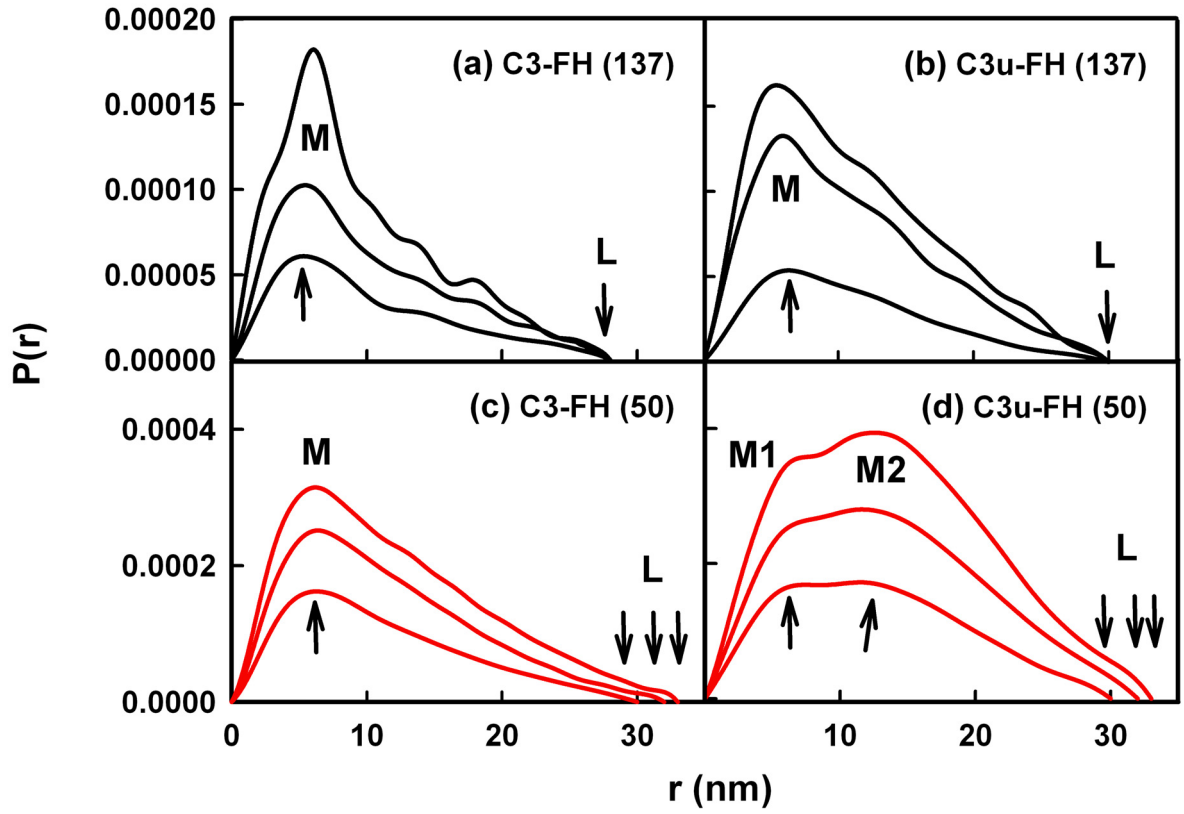


Figure 6.8 X-ray distance distribution functions *P(r)* for 1:1 mixtures of C3-FH and C3u-FH.

(a) For C3-FH between 0.3 mg/ml to 0.78 mg/ml in 137 mM NaCl, the peak maximum *M* is observed between 5.5 and 6.2 nm, and the maximum length *L* is 28 nm.

(b) For C3u-FH between 0.35 mg/ml to 0.93 mg/ml in 137 mM NaCl, *M* is observed at 5.5 to 6.4 nm and *L* is 30 nm.

(c) For C3-FH in 50 mM NaCl, *M* is observed at 6.3 nm, and *L* is 30 nm at 0.66 mg/ml, 32 nm at 0.99 mg/ml, and 33 nm at 1.32 mg/ml.

(d) For C3u-FH in 50 mM NaCl, the first peak *M1* is observed at 6.3 nm (0.52 mg/ml), 6.3 nm (0.78 mg/ml) and 6.3 nm (1.04 mg/ml), the second peak *M2* is observed at 11.7 nm (0.52 mg/ml), 12.2 nm (0.78 mg/ml) and 12.9 nm (1.04 mg/ml), and *L* is 30 nm (0.52 mg/ml), 32 nm (0.78 mg/ml) and 33 nm (1.04 mg/ml).

Table 6.1 Experimental sedimentation coefficient and X-ray scattering parameters for C3u-FH mixtures.

	FH Peak 1 (S)	%	C3u Peak 2 (S)	%	1:1 complex Peak 3 (S)	%	2:1 complex Peak 4 (S)	%	R_G (nm) [a]	D_{max} (nm)
C3-FH (137 mM NaCl)	5.91 ± 0.06	52 ± 4 %	8.21 ± 0.17	48 ± 4 %	n.o.		n.o.		7.83 ± 0.58 8.12 ± 0.58	28
C3u-FH (137 mM NaCl)	5.78 ± 0.09	44 ± 4 %	7.63 ± 0.20	6 ± 4 %	9.62 ± 0.32	50 ± 8 %	n.o.		8.83 ± 0.34 8.80 ± 0.29	30
C3-FH (50 mM NaCl)	5.80 ± 0.08	47 ± 4 %	7.97 ± 0.10	53 ± 4 %	n.o.		n.o.		8.07 ± 1.22 8.71 ± 0.62	30-33
C3u-FH (50 mM NaCl)	5.88 ± 0.07	31 ± 5 %	7.90 ± 0.40	8 ± 2 %	9.65 ± 0.45	51 ± 8 %	12.47 ± 0.70	10 ± 5 %	9.66 ± 1.16 9.90 ± 0.99	30-33

n.o., not observed

[a] The first value of each pair is from the Guinier R_G analyses at zero concentration (Figures 6.5 and 6.6) and the second is from the GNOM *P(r)* analyses (Figure 6.8).

[b] Peak 1 correspond to free FH, Peak 2 corresponds to free C3u, Peak 3 corresponds to the 1:1 C3u-FH complex, and Peak 4 corresponds to a 2:1 complex, being the C3u-FH-C3u complex and/or the $C3u_2$-FH complex.

C3u has an *M* value of 5.0 nm (Li *et al*., 2010; Chapter 5). Even though no change in *L* and *M* values were seen between the free proteins and their mixture, the appearance of the overall *P(r)* curve is different from that for the C3-FH mixture in Figure 6.8(a). In combination with the AUC data of Figure 6.3(d), the *P(r)* curves indicated that C3u and FH interact with each other in 137 mM NaCl, and there is no change in the overall length of the complex.

(iii) In 50 mM NaCl, the *L* values of the C3-FH mixture increased from 30 nm at the lowest concentrations of 0.66 mg/ml up to 33 nm at 1.32 mg/ml. The C3-FH mixture showed an *M* value of 6.3 nm (Figure 6.8(c)). In 50 mM NaCl, unbound C3 and FH have *M* values of 5.0 to 5.4 nm, and 5.4 nm respectively. The changes seen with these *P(r)* curves are best explained by the self-association of C3 and FH in low salt conditions (Li *et al*., 2010; Chapter 5; Okemefuna *et al.,* 2009a).

(iv) In 50 mM NaCl, the *L* values of the C3u-FH mixture increased from 30 nm at the lowest concentrations of 0.52 mg/ml up to 33 nm at 1.04 mg/ml (Figure 6.8(d)). Here, the overall appearance of the *P(r)* curve is different compared to the others in Figure 6.8. The C3u-FH mixture showed two *M* values, with an *M1* value at 6.3 nm and a second *M2* value at about 11.7 nm. In 50 mM NaCl, C3u and FH have different *M* values of 5.0 to 5.2 nm and 5.4 nm respectively. The presence of these large changes showed that further C3u-FH complexes have formed as the result of reduced salt concentration, in agreement with the AUC data of Figure 6.3(h). Interestingly, there are again no indications of changes in the length of FH in this complex.

Previous literature showed that SCR-1/4 and SCR-19/20 in FH possess binding sites for C3b (Sharma & Pangburn, 1996; Jokiranta *et al*., 2000; Hellwage *et al*., 2002; Schmidt *et al*., 2008). The cartoon to scale of C3u and FH (Figure 6.1) shows that these two C3b binding sites are well separated in FH. For both to interact with one molecule of C3u simultaneously, it would be necessary for FH to form a more compact structure. The unchanged or increased *L* values in the *P(r)* curves of Figure 6.8(b,d) do not indicate that such conformational changes take place. The *P(r)* curves are more in accord with the independent binding of two C3u molecules to separate sites on one FH molecule.

(6.2.4) Constrained modelling of C3-FH and C3u-FH mixtures

The modelling of the $s_{20,w}$ values and X-ray scattering curves for the C3-FH and C3u-FH mixtures is complicated by the weak interactions and multiple components

present in these mixtures. By this study, the availability of AUC and SAXS data sets for both C3 and C3u in 50 mM NaCl and 137 mM NaCl at several molar ratios with FH provides experimental information that enabled the interaction of C3u with FH to be unravelled. This modelling was based on earlier solution structures for free C3u (Figure 6.9(b)) and FH (PDB codes 3MMQ, 3GAV) (Li *et al*., 2010, Okemefuna *et al*., 2009a). A newer FH model was recently generated using data for homozygous FH in 137 mM NaCl and structures for five major FH fragments (Figure 6.9(a)) (Nan *et al.,* 2010; Okemefuna *et al*., 2010a). The crystal structure of the C3b complex with FH SCR-1/4 (PDB code 2WII) (Wu *et al*., 2009) enabled the superimposition of a full-length FH structure onto C3b to create a 1:1 C3u-FH complex at SCR-1/4 for curve fitting (Figure 6.9(d)). The AUC data showed that a 2:1 complex of C3u formed with FH. One possible structure corresponding to a $C3u_2$-FH complex was created by superimposition of the C3b dimer crystal structure onto the 1:1 C3u-FH model at SCR-1/4 (Figure 6.9(c)). Another possible structure was formed by two C3u molecules that interact independently with FH binding sites at SCR-1/4 and SCR-19/20. A model for two C3u molecules bound to FH was arbitrarily created at SCR-19/20 by recourse to biochemical data showing the residues involved in the interaction between the TED domain (C3d) and SCR-19/20 (Figure 6.9(e); Table 6.2) (Lambris *et al*., 1988; Sánchez-Corral *et al.,* 2002; Hellwage *et al.,* 2002; Manuelian *et al.,* 2003; Jozsi *et al.,* 2006; Lehtinen *et al.,* 2009; Jokiranta *et al.,* 2006).

The AUC and SAXS modelling of the C3u-FH mixtures followed the modelling procedure for pentameric-decameric mixtures of C-reactive protein based on known structures for the pentamer and decamer (Okemefuna *et al.,* 2010b). That previous study was validated by the similarity of the resulting three K_D values from three analyses of the pentamer-decamer association. Here, the six models for FH, C3, C3u, the 1:1 C3u-FH complex and two possible $C3u_2$-FH and C3u-FH-C3u complexes were first evaluated using their predicted $s^0_{20,w}$ values (Tables 6.1 and 6.3(a)). The difference in $s^0_{20,w}$ values for FH was within 0.3-0.8 S; that for C3 was within 0.1 S; that for C3u was within 0.4 S; that for the 1:1 C3u-FH complex was within 0.3 S; and that for two possible C3u-FH-C3u or $C3u_2$-FH complexes was within 0.4 S or within 0.8 S respectively. An alternative model for the 1:1 C3u-FH complex was formed from C3u bound to FH at SCR-19/20. This gave predicted $s^0_{20,w}$ values of 8.9-9.2 S, which was within 0.7-0.4 S of the experimental value of 9.6 S (Table 6.3(a)). For antibody and complement proteins, the agreement between predicted and experimental values is on

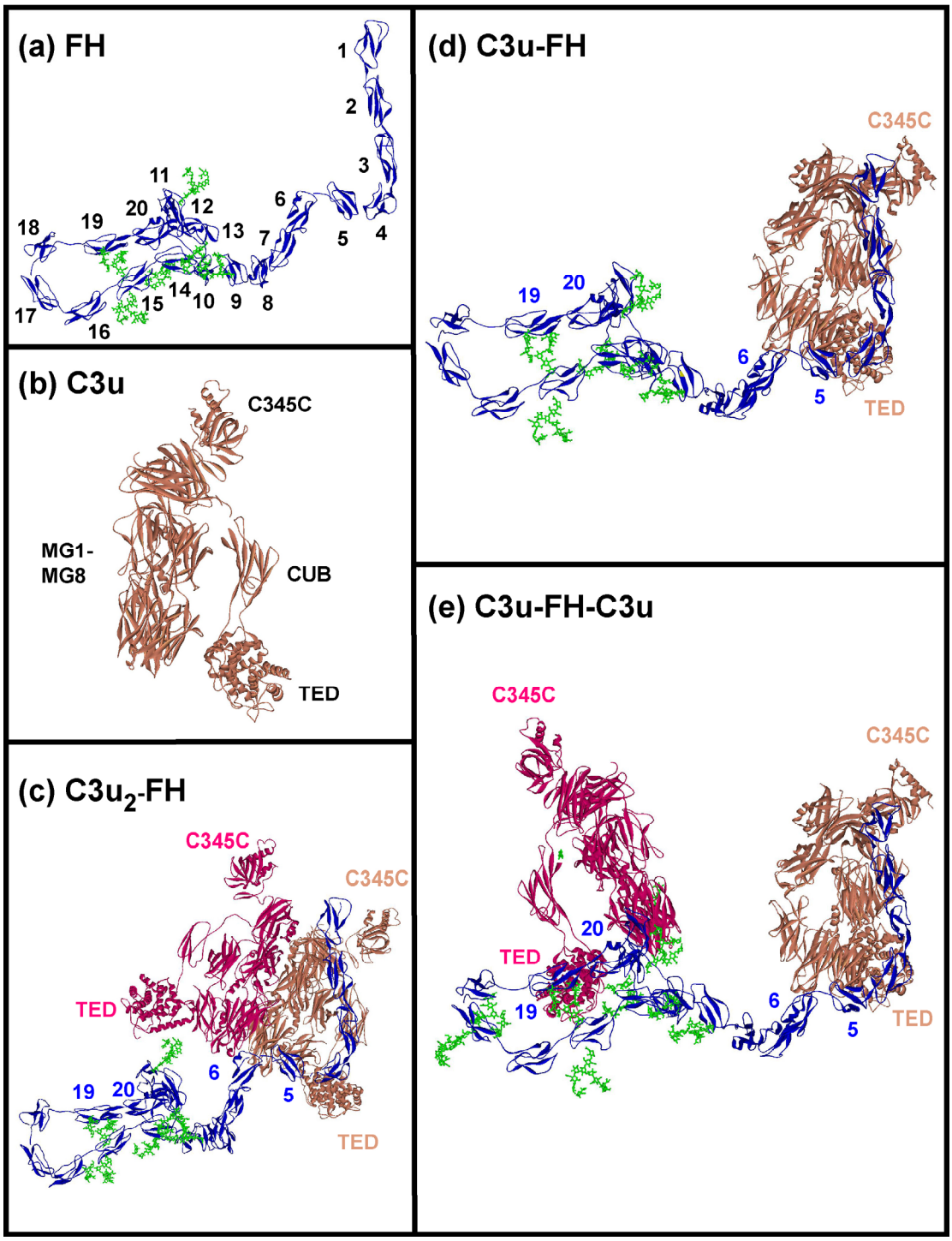


Figure 6.9 Molecular views of the FH, C3u and C3u-FH structures used for X-ray scattering curve fits.

(a) The 20 SCR domains of the solution structure for FH (PDB code 3N8O) are coloured in blue, with eight carbohydrate chains shown in green. The 20 domains are numbered as shown. The cartoon of Figure 6.1(b) corresponds to this view.

(b) The C3u solution structure (PDB code 3MMQ) is shown in tangerine, with the TED, CUB, C345C and MG1-MG8 domains labelled as shown.

(c) The $C3u_2$-FH complex is represented by the superimposition of the solution structure for FH with the crystal structure of the FH SCR-1/4 complex with C3b and the crystal structure of the C3b dimer (PDB codes 3N8O, 2WII and 2WIN).

(d) The C3u-FH complex is represented by the superimposition of the solution structure for FH with the crystal structure of the FH SCR-1/4 complex with C3b (PDB codes 3N8O and 2WII).

(e) The C3u-FH-C3u complex is represented by the superimposition of the solution structure for FH with the crystal structure of the FH SCR-1/4 complex with C3b and the C3u solution structure in an arbitrary orientation at SCR-19/20 (PDB codes 3N8O, 2WII and 3MMQ).

Table 6.2 Binding sites of FH SCR-19/20 with C3b and C3

SCR-19/20 residues	C3b (C3d) residues	Method	References
Q1139, W1157, R1182, W1183, K1188, R1206, R1210, T1184		C3b mutagenesis, radioligand assay	(Lehtinen *et al*., 2009)
R1182, W1183, T1184, K1188, E1198, R1215		C3b mutagenesis, SPR and docking	(Jokiranta *et al*., 2006)
W1183, V1197, R1210		C3b ELISA	(Sanchez-Corral *et al*., 2002)
E1172, R1210, R1215		C3d SPR	(Manuelian *et al*., 2003)
W1157, W1183, V1197, R1210, R1215, P1226		C3b and C3d ELISA,SPR	(Józsi *et al*., 2006)
	K1187- A1249	C3b mutagenesis, ELISA and radioligand assay	(Lambris *et al*., 1988)
R1171, E1172, M1174, E1175, N1178, L1181, R1182, K1202, E1224, Y1225, P1226, R1231	D1216(D223), N1218(N225), R1219(R226), E1221(E228), Q1226(Q233), D1247(D254), P1251(P258), R1254(R261), W1255(W262), E1258(E265), Q1259(Q266), R1260(R267) [a]	C3d docking	(Hellwage *et al*., 2002)

[a] The first residue number corresponds to C3b (PDB code 2I07), and the parenthesized residue numbering corresponds to C3d.

[b] The thioester covalent surface-attachment site on C3d is about 3 nm away from the FH binding site in the SCR-1/4 complex with C3b.

Table 6.3 X-ray scattering and sedimentation coefficient modelling fits for the solution structures of C3u and FH

(a) Single models	Models	Hydrated spheres	R_G (nm)	D_{max} (nm)	R-factor (%)	Calculated $s^0_{20,w}$ (S)	Experimental $s^0_{20,w}$ (S)				
FH	1	2126	8.29	29	8[a], 9.4[b]	5.1, 5.5	5.77-5.91				
C3	1	1863	4.35	16	16[a]	8.11	7.97-8.21				
C3u	1	1877	4.78	19	13.6[b]	7.48	7.63-7.90				
C3u-FH	1	1433	9.11	31	8.5[b]	9.38; 8.85-9.20	9.62-9.65				
C3u-FH-C3u	1	2104	10.70	33	22.6[b]	12.87	12.47				
$(C3u)_2$-FH	1	2033	9.12	30	17.9[b]	13.30	12.47				
(b) C3/C3u-FH modelling						FH (%)	C3(%)	C3u(%)	C3u-FH (%)	$(C3u)_2$FH (%)	C3u-FH-C3u (%)
C3-FH (137 mM NaCl)	101		4.89 - 8.17	n.a	3.9 - 15.9						
	best fit		7.25	26	3.9	72	28	-	-	-	-
C3u-FH (137 mM NaCl)	1326		5.41 - 9.11	n.a	3.6 - 13.6						
	best fit		8.59	28	3.6	46	-	2	52	-	-
C3-FH (50 mM NaCl)	101		4.89 - 8.17	n.a	10.7 – 17.0						
	best fit		6.63	26	10.7	53	47	-	-	-	-
C3u-FH (50 mM NaCl)	316,260		5.41 - 10.71	n.a	8.5 - 27.5						
	best fit		9.23	30	8.5	20	-	0	0	60	20

n.a., Not available

[a] Compared with the C3-FH data in 137 mM NaCl
[b] Compared with the C3u-FH data in 137 mM NaCl
[c] The first value corresponds to C3u-FH complexed at SCR-1/4; the second to that complexed at SCR-19/20.

average within ± 0.21 S, and not normally greater than 0.3 S (Perkins *et al.,* 2009). It was concluded that the six predicted and experimental $s^0_{20,w}$ values were in satisfactory agreement with each other (Table 6.3(a)). For the complexes of this study, the $s^0_{20,w}$ differences can be larger because of reduced precisions in determining the experimental values in multicomponent mixtures, especially for the broad peak 4. These agreements show that these structural models could be used for fits to the SAXS curves. Several conclusions were drawn from these modelling results: (i) The explanation of the $s^0_{20,w}$ value for the 1:1 C3u-FH complex required no conformational change in FH; (ii) It was not possible to show which of the 1:1 complexes at SCR-1/4 or SCR-19/20 was preferred; (iii) The calculations showed a preference for a C3u-FH-C3u complex over a $C3u_2$-FH complex in order to explain peak 4 of Figure 6.3; (iv) All the $s^0_{20,w}$ values were explained by the existence of well-defined complexes with no reaction boundaries present.

The SAXS curve fits for the non-interacting C3-FH mixtures were based on comparisons of the curves calculated from the C3 and FH models with the experimental X-ray curves in 137 mM NaCl and 50 mM NaCl. The experimental R_G values of 7.83-8.07 nm for C3-FH in both buffers (Table 6.1) were similar to the modelled R_G value of 8.29 nm for FH alone, but less so to that of 4.35 nm for C3 alone (Table 6.3). Visually, the modelled scattering curves for FH alone (blue line) and C3 alone (red dash) do not match the C3-FH experimental curve in 137 mM NaCl (cyan circles) with poor high R-factors (goodness-of-fit) of 8% and 16% respectively (Figure 6.10(a); Table 6.3). Likewise the modelled FH and C3 curves do not match the C3-FH experimental curve in 50 mM NaCl with even poorer higher R-factors of 15% and 18.6% (Figure 6.10(b)). A total of 101 combinations of the C3 and FH modelled curves in increments of 1% was created. The 101 *R*-factors ranged between 3.9-15.9% and 10.7–17.0 % for 137 mM NaCl and 50 mM NaCl respectively, for R_G values that ranged between 4.89-8.17 nm (Figure 6.11(a,c)). For 137 mM NaCl, the lowest R-factor of 3.9% corresponded to a mixture of 28% C3 and 72% FH. This combination showed a good visual agreement with the experimental curve for C3-FH in 137 mM NaCl (Figure 6.10(e)). The proportions of C3 and FH were comparable with the SEDFIT integration of the *c(s)* plots that showed 48% C3 and 52% FH in 137 mM NaCl (Figure 6.3(b); Table 6.1). For 50 mM NaCl, the lowest R-factor of 10.7% corresponded to a mixture of 47% C3 and 53% FH (Table 6.3). This combination showed a fair visual agreement with the experimental curve for C3-FH in 50 mM NaCl, bearing in mind that the experimental

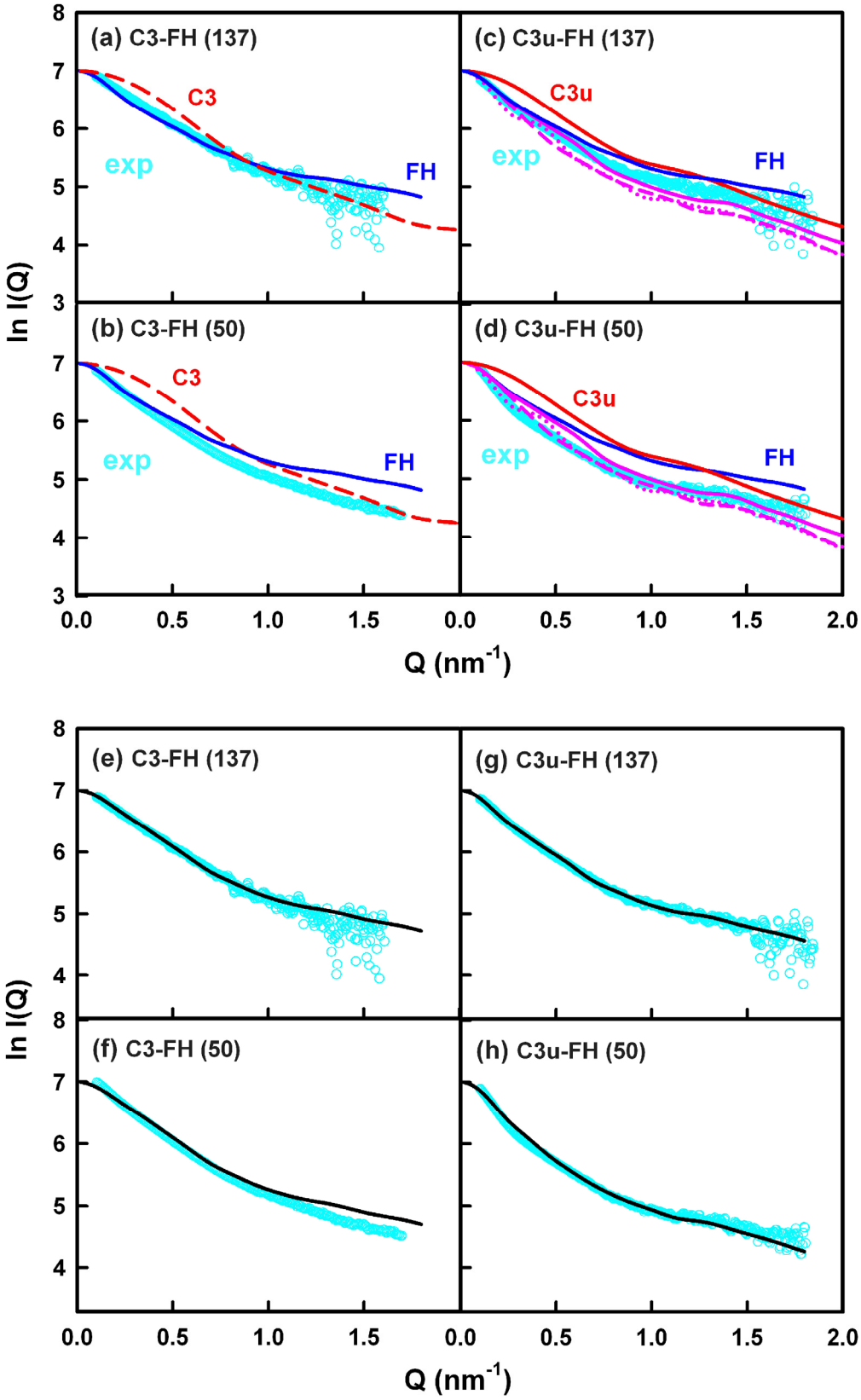


Figure 6.10 Comparison of the experimental and modelled X-ray scattering curves (Legend overleaf).

Figure 6.10 Comparison of the experimental and modelled X-ray scattering curves. The experimental data are shown in cyan. The modelled curves are shown as lines (C3, red dashed; C3u, red solid; FH, blue).

(a,b) For the C3-FH mixture in 50 mM and 137 mM NaCl buffer, the experimental curve is compared with the modelled curves for C3 and FH.

(c,d) For the C3u-FH mixture in 50 mM and 137 mM NaCl buffer, the experimental curve is compared with five modelled curves for C3u, FH, and the C3u-FH, $C3u_2$-FH and C3u-FH-C3u complexes (Figure 6.9). (C3u-FH, pink; $C3u_2$-FH, pink dashed; C3u-FH-C3u, pink dotted).

(e,f) The best-fit combination of two curves for the analyses in C3-FH in 137 mM and 50 mM NaCl. Here and in (g,h), the black line represents the best-fit calculated curve that was summed from the combination of separate curves.

(g) The best-fit combination of the three curves for C3u, FH and the 1:1 C3u-FH complex in 137 mM NaCl.

(h) The best-fit combination of the five curves for C3u, FH, and the 1:1 and 2:1 C3u-FH complexes in 50 mM NaCl.

curve showed evidence of protein aggregation (Figure 6.10(f)). The proportions also agreed well with the SEDFIT integration of the *c(s)* plots that showed 53 ± 4% C3 and 47 ± 4% FH (Table 6.1). The agreement between the SAXS- and AUC-determined proportions of C3 and FH justified the applicability of the X-ray curve fit method to analyse the interactions between C3u and FH.

Next, the SAXS fits for the C3u-FH interactions were evaluated. Greater complexity resulted because a weak interaction takes place and the unbound components as well as their complexes need consideration in the modelling. For this, the five models for FH, C3u, the 1:1 C3u-FH complex and two possible $C3u_2$-FH and C3u-FH-C3u complexes were used.

(i) For the C3u-FH mixture in 137 mM NaCl, the experimental R_G value was 8.83 ± 0.34 nm (Table 6.1). The modelled curves for C3u, FH, and the three complexes all visibly deviated from the experimental *I(Q)* curve with high R-factors of 13.6% (C3u), 9.4% (FH), 8.5% (the 1:1 C3u-FH complex), 22.6% (the C3u-FH-C3u complex) and 17.9% (the $C3u_2$-FH complex) (Figure 6.10(c)). A total of 1326 different combinations of the summed C3u, FH and the 1:1 C3u-FH curves in 2% increments showed that the resulting *R*-factors ranged satisfactorily between 3.6-13.6%, and the R_G values ranged satisfactorily between 5.41-9.11 nm (Figure 6.11(b); Table 6.3(b)). The experimental R_G value of 8.83 nm is close to the predicted R_G value of 8.59 nm at the minimum *R*-factor of 3.6%. This combination corresponded to a composition of 2% C3u, 46% FH, and 52% of the 1:1 C3u-FH complex (Table 6.3(b)). The X-ray-fitted proportions agreed well with the SEDFIT integration of the *c(s)* plots that showed 6% C3u, 44% FH, and 50% C3u-FH (Figure 6.3(d); Table 6.1). By converting the percentage of C3u, FH and 1:1 complex to molarities, this composition resulted in a K_D value of 0.22 ± 0.06 μM. This is comparable with the K_D determination of 0.59 ± 0.11 μM from the *c(s)* analysis of Figure 6.3(d). The use of all five components in 2% increments for fitting did not result in an improved curve fit (not shown), although here the best-fit combination was comparable at 5% C3u, 50% FH, 25% C3u-FH, 0% C3u-FH-C3u and 20% $C3u_2$-FH. Overall, we conclude that an analysis of C3u, FH and the 1:1 C3u-FH complex presented the simplest explanation for the observed X-ray curve in 137 mM NaCl. This combination resulted in an excellent visual agreement with the experimental curve for this C3u-FH mixture (Figure 6.10(g)).

(ii) For the C3u-FH mixture in 50 mM NaCl, the experimental R_G value was 9.66 ± 1.16 nm (Table 6.1). The modelled curves for C3u, FH, and the three complexes

all visibly deviated from the experimental *I(Q)* curve again. The experimental curve now resembled the curves for C3u-FH-C3u and $C3u_2$-FH the most closely, in reflection of the larger proportion of complexes formed in low salt (Figure 6.10(d)). The testing of 316,260 different combinations of the five curves gives the best fit with proportions of 0% C3u, 20% FH, 0% C3u-FH, 20% C3u-FH-C3u, and 60% $C3u_2$-FH. The R_G value of 9.23 nm at the lowest R-factor of 8.5% from the best-fit model is smaller than the experimental R_G value of 9.66 nm (Figure 6.11(d); Table 6.3(b)). It is comparable with the AUC integration that gave 8% of C3u, 31% of FH, 51% of the 1:1 complex and 10% of a 2:1 complex (Table 6.1). While the percentages of 1:1 and 2:1 complexes are different between the X-ray and AUC integrations, their totals were 80% and 61% respectively, showing that large amounts of FH complexes with one or two C3u molecules were present in 50 mM NaCl buffer. The most likely reason for the difference between the SAXS and AUC integrations is the need for a good structure for the 2:1 complex in order to fit this to the SAXS curves. The best-fit C3u-FH combination resulted in a very good visual agreement with the experimental curve for the C3u-FH mixture in 50 mM NaCl (Figure 6.10(h)).

The availability of SAXS curves with other molar ratios of C3u and FH (Figure 6.7) will make possible a more extensive analysis of the molar ratios of unbound and complexed proteins. This will provide K_D values for complex formation for comparison with those from AUC.

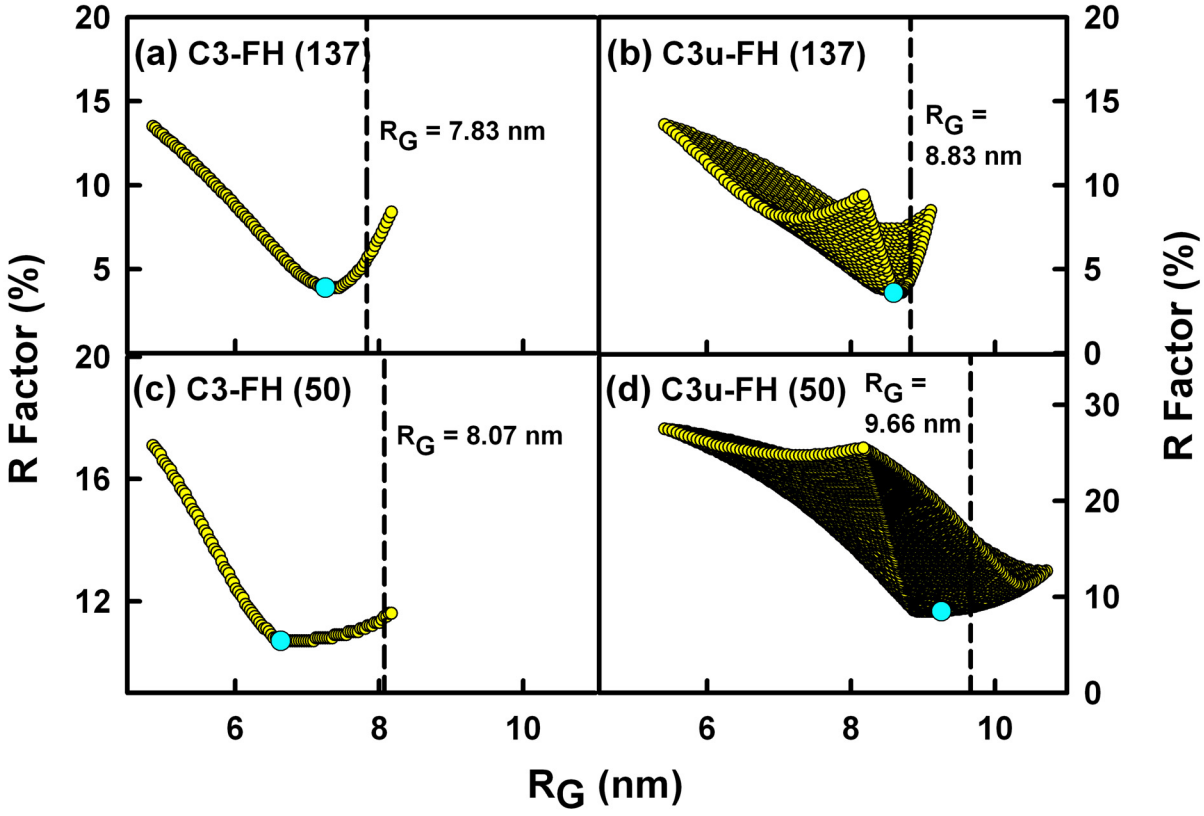


Figure 6.11 Modelling analyses for four C3-FH and C3u-FH mixtures. The vertical dashed lines correspond to the measured experimental R_G values. All the trial combinations of curves are shown in yellow, while the best-fit curve is shown in cyan (Table 6.3).

(a,c) The R-factor values from 100 combinations of the C3 and FH modelled curves in 1% increments are compared with their R_G values.

(b) The R-factor values for 1326 combinations of the three C3u, FH and C3u-FH modelled curves in 2% increments are compared with their R_G values.

(d) The R-factor values for 316,260 trial combinations of the five C3u, FH, C3u-FH, $C3u_2$-FH and C3u-FH-C3u modelled curves in 2% increments are compared with their R_G values.

(6.3) Conclusions

The novel conclusion of this study is that two C3u molecules are able to bind independently at separate sites in FH to a relatively inflexible FH structure in solution. The "tick-over" conversion of C3 to C3u leads to alternative pathway activation, because C3u exhibits C3b-like function and is able to form an initial convertase with the Bb fragment of Factor B to enable the rapid generation of active C3b (Law & Reid, 1995). The interaction of the major complement regulator FH with C3u offers regulatory control of the initial activation process, and provides insight into the similar processes involving FH regulatory control of fully activated C3b. There are two well-established C3b binding sites at FH SCR-1/4 and FH SCR-19/20 (Sharma & Pangburn, 1996; Jokiranta *et al*., 2000; Hellwage *et al*., 2002; Schmidt *et al*., 2008) (Table 6.2). Because of the knowledge of two C3b sites in FH, it has often been presumed that the 20 SCR domains in FH are joined by flexible linkers that are able to bend in order that the two sites are brought together to enable FH to bind to a single C3b molecule, or alternatively FH binds to two distinct C3b molecules bound to the surface of the pathogen (Herbert *et al*., 2006; Jokiranta *et al*., 2006; Schmidt *et al*., 2008). Here, we have performed experiments with (i) both C3u and C3 (with C3 acting as a non-binding control), and (ii) both physiological 137 mM NaCl and reduced salt 50 mM NaCl (with low salt acting as a promoter for the FH-C3u interaction, this being a further control of the FH-C3u interaction). There is no evidence from the $s_{20,w}$ values for any major conformational changes in FH in the presence of C3u. This indicates that its overall structure is not flexible. It was also noticeable that the proportion of C3u (peak 2) in equimolar FH-C3u mixtures was diminished, in distinction to the more equal proportions of C3 and FH in C3-FH mixtures. Furthermore, the formation of a previously-unreported 2:1 complex (peak 4) in addition to the 1:1 complex (peak 3) indicated that two C3u molecules were bound to one FH molecule. Hence this study has clarified the structures and binding stoichiometry involved in the FH-C3u interaction at physiological NaCl concentrations.

The extent to which the presents results are applicable to the C3b-FH interaction depend on how similar C3u and C3b are. C3u is prepared by the hydrolysis of the thioester in the TED domain using hydrazine, making this easier to prepare in large amounts than C3b but retains the C3a (ANA) fragment (Figure 6.1). C3b is purified by tryptic cleavage of C3 and has lost the ANA domain, but retains an intact thioester. The comparison of the C3u solution structure with the crystal structures of C3b and C3

shows that C3u resembles C3b more closely than C3 (Li *et al.,* 2010; Chapter 5). The similarity of the C3u-FH and C3b-FH interactions is further indicated by the similar K_D of 0.59 μM for both complexes (Table 6.4). C3u forms dimers, and C3b dimers have been seen in one crystal structure (Li *et al.,* 2010; Rooijakkers *et al.*, 2009). It is likely that C3u binds to FH SCR-1/4 in a similar way to that of C3b with SCR-1/4 in the crystal structure of this complex (Wu *et al.* 2009). Nonetheless, further similar studies of C3b-FH mixtures will be required to complete this project.

FH enables proteolytic degradation of C3u and C3b by providing a binding platform for the serine protease Factor I while stabilizing the overall domain arrangement of C3b. The latter is achieved by the cross-linking of the TED and MG1 domains by the SCR-4 domain (Wu *et al.* 2009). In unbound C3u, the CUB and TED domains are in an extended and flexible arrangement away from its MG domains, this being mediated by the linker peptides of length about 9 to 10 residues (Li *et al*., 2010). Flexibility has also been demonstrated by electron microscopy that showed that the CUB and TED move freely in C3u (Nishida *et al*., 2006). In distinction, the lack of structural change in FH indicates that the inter-SCR linkers are less flexible than commonly thought. In the past, the presumption of flexibility in the inter-SCR linkers in FH that enable different SCR arrangements to be formed came from several sources. First, crystal structures for the SCR proteins do not show uniformity in their inter-SCR orientations, and it has not proved possible to predict their orientations (Perkins *et al.,* 2002). Second, the FH linkers are of lengths 3-8 residues, where six of the seven longest linkers are found between SCR-7 and SCR-15 where they might cause the centre of FH to become mobile. This is suggested by analogy with the 8-residue linker between SCR-1 and SCR-2 in complement receptor type 2, which resulted in a compact folded-back domain arrangement in its crystal structure, but was opened up into an extended structure in solution (Szakonyi *et al.,* 2001; Gilbert *et al.,* 2005; Prota *et al.,* 2002). The lack of structural flexibility in FH when interacting with C3u may be related to the eight oligosaccharides chains at SCR-9 and between SCR-12 to SCR-18 (Fenaille *et al.,* 2007) that may act to separate the N-terminal and C-terminal domains into two functionally distinct regions.

The existence of two separate binding sites for C3u on FH can be related to two different functionalities in FH. The binding site at SCR-1/4 relates more to the regulatory interaction between FH and C3u/C3b that leads to its proteolytic cleavage

Table 6.4 Dissociation constants K_D for the FH-C3b interaction.

FH	Ligand	Buffer	Techniques	Experimental K_D (μM)	Temperature	References
SCR-1/4	C3b	10 mM sodium phosphate, 150 mM NaCl, 0.005% Tween-20, pH 7.4	SPR	11 ± 2	25°C	(Wu *et al*., 2009)
SCR-1/4	C3b	10 mM HEPES-buffered, 150 mM saline, 3 mM EDTA, 0.05% (v/v) surfactant p20, pH 7.4.	SPR	9.8±0.3[a] to 13.5± 0.2[b]	25°C	(Schmidt *et al*., 2008)
SCR-19/20	C3b	10 mM HEPES-buffered, 150 mM saline, 3 mM EDTA, 0.05% (v/v) surfactant p20, pH 7.4.	SPR	3.5 ± 0.1[b] to 4.5 ± 0.5[a]	25°C	(Schmidt *et al*., 2008)
SCR-19/20	C3b	10 mM acetate buffer (pH 5.0 –5.7)	SPR	5.4 ± 0.7[a]	22°C	(Jokiranta *et al*., 2006)
SCR-1/20	C3b	10 mM HEPES-buffered, 150 mM saline, 3 mM EDTA, 0.05% (v/v) surfactant p20, pH 7.4.	SPR	0.59 ± 0.04[a] to 1.6 ± 0.1[b]	25°C	(Schmidt *et al*., 2008)
SCR-1/20	C3u	8.1 mM Na_2HPO4, 1.5 mM KH_2PO4, 2.7 mM KCl, 50 mM NaCl, pH 7.4	AUC	0.11 ± 0.06	20°C	This present study
SCR-1/20	C3u	8.1 mM Na_2HPO4, 1.5 mM KH_2PO4, 2.7 mM KCl, 137 mM NaCl, pH 7.4	AUC	0.59 ± 0.11	20°C	This present study
SCR 16-20	C3d (TED of C3b)	10 mM HEPES, 50 mM NaCl, pH 7.4	AUC	1.6	20°C	(Okemefuna *et al*., 2009b)
SCR 16-20	C3d (TED of C3b)	10 mM phosphate, 137 mM NaCl, pH 7.4	AUC	29	20°C	(Okemefuna *et al*., 2009b)

[a] Results from C1 sensor chips

[b] Results from CM5 sensor chips

and C3u/C3b degradation. The binding site at SCR-19/20 binds to C3d as well as to C3b/C3u, and the C3d interaction leads to multimerisation events that will promote the binding of more FH to surfaces that have high amounts of C3d bound to them (Okemefuna *et al.,* 2009b). Such surfaces correspond to host cells that were decorated with bound C3b which were subsequently degraded to C3d. This same FH SCR-19/20 region also binds to heparin and C-reactive protein, both of which will decorate host cells under appropriate conditions (Okemefuna *et al*., 2010c). Similar binding functionality also exists at SCR-7 which also binds to heparin and C-reactive protein. In other words, the binding of SCR-19/20 to C3u, C3b or C3d is more relevant in a context of the host cell surface, rather than providing a synergistic means of enhancing C3b breakdown mechanisms.

Studies of the interactions between the complement proteins are dominated by their weak interactions, in which the μM K_D values mean that 100% complex formation will never be achieved in a physiological scenario. This was confirmed directly for the C3u-FH mixtures in which the AUC- and SAXS-measured sample compositions were typically 2-6% for C3u, 44-46% for FH and 50-52% for the 1:1 complex in 137 mM NaCl using physiological μM concentration values (Tables 6.1 and 6.3). The use of reduced non-physiological NaCl concentrations increased the proportions of the complexes, showing that the interaction is electrostatic. The compilation of K_D values (Table 6.4) shows that the interaction between SCR-1/4 and C3b and that between SCR-19/20 have K_D values that are about 10 times greater than that for intact FH interacting with C3b. This discrepancy is most easily explained by the assumption of a 1:1 interaction model for the C3b-FH study in previous studies (Schmidt *et al*., 2008). If this is replaced by a 2:1 interaction model for the C3b-FH complex, the joint summation of the two separate binding sites at SCR-1/4 and at SCR-19/20 wil lead to a much reduced K_D value for full-length FH. It should be noted that the AUC and SAXS data provide an independent method for determining associations compared to those in earlier studies (Table 6.4), because both methods correspond to a fluid phase measurement, not one based on binding to a sensor chip surface as with surface plasmon resonance.

A side-benefit of this study has been the successful demonstration of a new method for determining K_D values of weakly interacting macromolecules in solution based on SAXS curve fits. Previously, one approach for K_D determinations is through

the use of surface plasmon resonance methods, in which a ligand is bound to a sensor surface and an analyte is flowed over this (Tanios *et al.,* 2008). By AUC methods, it is possible to determine K_D values, either by exponential curve fit analyses of sedimentation equilibrium data, or by integration of the peaks in the *c(s)* size distribution plots from analyses of sedimentation velocity data (Cole *et al.,* 2008). By X-ray scattering, provided that scattering curves are known for the individual non-interacting macromolecules A and B and their complex C, and provided that no large shape changes occur during complex formation, it is possible to determine the ratio of the unbound and bound species by the optimisation of the best-fit combination of the scattering curves for A, B and C. This was recently illustrated for C-reactive protein, where combinations of the scattering curves for its pentamer and decamer forms resulted in a K_D value of 16 μM, in good agreement with two other K_D determinations in that study (Okemefuna *et al.,* 2010b). Here, the application of this SAXS fit approach showed that it was possible to generate approximately similar proportions of C3 and FH for mixtures of C3-FH in 137 mM and 50 mM NaCl that agreed well with the *c(s)* integrations from AUC data (Tables 6.1 and 6.3). By the same token, the similar proportions of 2% C3u, 46% FH, and 52% from SAXS fits and 6% C3u, 44% FH, and 50% C3u-FH by AUC integrations indicate that this SAXS approach has promise for determining K_D values.

(6.4) Material and Methods

(6.4.1) Preparation of C3-FH and C3u-FH mixtures

The purification of C3 from fresh human plasma and C3u by incubation with hydrazine and its separation from unhydrolysed C3 by anion-exchange chromatography followed that described previously (Dodds, 1993; Li *et al.,* 2010). Reduced SDS-PAGE confirmed the presence of α chain (115 kDa) and β chain (75 kDa). C3u but not C3 was active in functional assays using factor I and factor H (Nan *et al.,* 2008). After confirmation of purity, the proteins were dialysed into phosphate buffer saline (PBS: 8.1 mM Na_2HPO4, 1.5 mM KH_2PO4, 2.7 mM KCl, 50 mM or 137 mM NaCl, pH 7.4) overnight at 4ºC, and concentrated using Amicon® Ultra-15 centrifugal filter devices with a molecular weight cut-off of 50 kDa at a speed of 2,500 g. Throughout the text, the buffer is denoted as 137 mM NaCl or 50 mM NaCl, even though phosphate is also present. Concentrations were determined using an absorption coefficient of 9.4 (1%, 280 nm, 1 cm path length) (Li *et al.,* 2010; Chapter 5). Prior to mixture with FH, C3 and C3u were gel filtrated to remove non-specific aggregates using a Superose 6 16/60 column (Amersham Pharmacia Biotech). FH was purified from a pool of anonymised human plasma using monoclonal affinity chromatography with MRC-OX23 Sepharose column (Sim *et al*., 1993; Nan *et al.,* 2008). Concentrations were determined using an absorption coefficient of 16.2 (Okemefuna *et al.,* 2009a). Non-specific aggregates and traces of human serum albumin were removed by gel filtration using a Superose 6 16/60 column (Amersham Pharmacia Biotech).

To prepare the mixtures of C3-FH and C3u-FH, the C3, C3u and FH proteins after gel filtration were dialysed individually into 137 mM NaCl or 50 mM NaCl phosphate buffers. The proteins were centrifuged to remove any visible aggregates, then mixed in a 1:1 ratio according to their absorbances at 280 nm. The mixtures were incubated at 20°C for 6 h with the addition of 1 mM Pefabloc protease inhibitor before loading onto a Superose 6 gel filtration column. For C3-FH and C3u-FH in 137 mM NaCl and 50 mM NaCl, the fractions at the peak centres were collected and concentrated for data collection. Prior to and after AUC and X-ray experiments, sample purities were checks by non-reduced and reduced SDS-PAGE (Figure 6.2(c)).

(6.4.2) Analytical ultracentrifugation of C3-FH and C3u-FH mixtures

Sedimentation velocity experiments were performed at 20°C using two Beckman XL-I analytical ultracentrifuges (Beckman-Coulter Inc, Palo Alto, CA) equipped with both absorbance and interference optics. An eight-hole AnTi50 rotor was

used with standard double-sector cells with column heights of 12 mm at 20ºC, and sedimentation was monitored using absorbance optics at 280 nm and interference optics. Below a total concentration of 1 mg/ml, both absorbance and interference optics data were analysed. Above 1 mg/ml, only interference optics data were used for reason of the saturation of the absorbance optics at 280 nm. Data for the C3-FH and C3u-FH mixtures were acquired in 137 mM and 50 mM NaCl phosphate buffer at rotor speeds of 30,000, 40,000 and 50,000 r.p.m. The C3-FH mixture in 137 mM NaCl used a total concentration of 1.43 mg/ml, prepared by mixing 320 µl of 1.26 mg/ml C3 and 260 µl of 1.72 mg/ml FH in a 1:1 molar ratio. The C3-FH mixture in 50 mM NaCl used a total concentration of 1.36 mg/ml, prepared by mixing 220 µl of 1.37 mg/ml C3 and 135 µl of 1.81 mg/ml FH in a 1:1 molar ratio. The C3u-FH mixture in 137 mM NaCl used a total concentration of 1.53 mg/ml, prepared by mixing 455 µl of 1.21 mg/ml C3u and 193 µl of 2.30 mg/ml FH in a 1:1 molar ratio. The C3u-FH mixture in 50 mM NaCl used a total concentration of 1.62 mg/ml, prepared by mixing 300 µl of 1.21 mg/ml C3u and 104 µl of 2.81 mg/ml FH in a 1:1 molar ratio. Each mixture was studied in a dilution series from 100% to 10%. The buffer density of 137 mM and 50 mM NaCl phosphate buffers were measured at 20^{o}C using an Anton-Paar DMA5000 density meter to be 1.00542 ml/g and 1.00167 ml/g, in good agreement with the predicted values of 1.00543 ml/g and 1.00175 ml/g from SEDNTERP (Laue *et al.,* 1992). A mean partial specific volume of 0.728 ml/g was used for the C3-FH and C3u-FH mixtures based on the values of 0.739 ml/g for C3 (mass 189 kDa) and 0.715 ml/g for FH (mass 154 kDa) calculated from its amino acid and carbohydrate content, respectively (Li *et al.,* 2010; Okemefuna *et al.,* 2009a). The continuous $c(s)$ size distribution analysis method was used to determine sedimentation coefficients $s_{20,w}$ of the C3-FH and C3u-FH mixtures using SEDFIT (version 11.7) (Dam & Schuck, 2004; Schuck, 2000). The final $c(s)$ fits were determined using a resolution of 200 and by floating the meniscus, frictional ratio f/f_0, and baseline and holding the cell bottom, partial specific volume and solvent density fixed until the overall root-mean-square deviations and visual appearance of the fits were satisfactory (Figure 6.3). Because f/f_0 is floated in the fits, the final f/f_0 values may be affected by oligomerisation. The SEDFIT analyses assumed that all species have the same frictional ratio *f/fo*. Molecular masses were determined by converting the *c(s)* plots to *c(M)* mass distributions within SEDFIT.

(6.4.3) X-ray scattering of C3-FH and C3u-FH mixtures

X-ray solution scattering data of the C3-FH and C3u-FH mixtures were collected

in four sessions on the camera at the ID02 high brilliance beamline at the European Synchrotron Radiation Facility (ESRF) at Grenoble, France with a ring energy of 6.0 GeV (Narayanan *et al*., 2001). In all four sessions except one, data were collected in 16-bunch mode using beam currents of 68-93 mA, 66-90 mA and 68-74 mA. In one session, data were collected in 4-bunch mode using beam currents of 32-34 mA. Data were acquired using an improved fibre optically-coupled high sensitivity and dynamic range CCD (charge-coupled device) detector (FReLoN) with a smaller beamstop. The sample-to-detector distance was 2.0 m. A flow cell was used to eliminate radiation damage in conjunction with on-line checks for the absence of radiation damage through the optimisation of exposure times in sets of ten time frames. Each sample was measured in four set of ten time frames of length 0.08, 0.1, 0.15, 0.2 and 0.3 sec for the C3-FH and C3u-FH mixtures. The time frames for each run were averaged to maximise signal-noise ratios. Sample temperatures corresponded to ambient conditions at 20°C. The C3-FH mixtures in 1:0.3 and 1:0.5 molar ratios in 137 mM NaCl were studied at four total concentrations of 0.32 mg/ml – 1.29 mg/ml and 0.32 mg/ml – 1.27 mg/ml respectively. The 1:1 C3-FH mixtures in 137 mM NaCl were studied at seven total concentrations of 0.30 mg/ml – 1.46 mg/ml. The C3-FH mixtures in 1:0.3 and 1:0.5 molar ratios in 50 mM NaCl were studied at single total concentrations of 0.56 mg/ml and 0.63 mg/ml respectively. The 1:1 C3-FH mixtures in 50 mM NaCl were studied at four total concentrations of 0.33 mg/ml – 1.32 mg/ml. C3u-FH 1:0.3 and 1:0.5 in PBS 137 mM NaCl were collected at 0.37 mg/ml, 0.75 mg/ml and 0.39 mg/ml, 0.80 mg/ml respectively. C3u-FH 1:1 in PBS, 137 mM NaCl were studied at 0.26 mg/ml – 1.04 mg/ml (nine concentrations). C3u-FH 1:0.3 and 1:0.5 in PBS 50 mM NaCl were collected at three concentrations from 0.40 mg/ml – 1.13 mg/ml and 0.42 mg/ml – 1.15 mg/ml respectively. C3u-FH 1:1 in PBS, 50 mM NaCl were studied at 0.22 mg/ml – 1.04 mg/ml (eleven concentrations).

Guinier analyses were used to characterise the mean degree of structural elongation of macromolecules in solution. Guinier plots at low Q values gives the R_G and forward scattering at zero angle I(0) (Glatter & Kratky, 1982):

$$\ln I(Q) = \ln I(0) - R_G^2 Q^2/3$$

For a single macromolecule, this expression is valid in a $Q.R_G$ range up to 1.5. In a given solute-solvent contrast, the radius of gyration R_G is a measure of structural elongation if the internal inhomogeneity within C3, C3u or FH has no effect. For elongated structures, the mean radius of gyration of the cross-sectional structure R_{XS} and

the mean cross-sectional intensity at zero angle $[I(Q).Q]_{Q\rightarrow 0}$ parameters are obtained from:

$$\ln [I(Q).Q] = \ln [I(Q).Q]_{Q\rightarrow 0} - R_{XS}^2 Q^2/2$$

The R_{XS} value monitors the elongation of its cross-section shape along its longest axis. The R_G and R_{XS} analyses were performed using an interactive PERL script program SCTPL7 (J. T. Eaton and S. J. Perkins, unpublished software) on Silicon Graphics OCTANE Workstations. Indirect Fourier transformation of the full scattering curve I(Q) in reciprocal space gives the distance distribution function $P(r)$ in real space. This yields the maximum dimension of the macromolecule L and its most commonly occurring distance vector M in real space:

$$P(r) = \frac{1}{2\pi^2} \int_o^\infty I(Q)\, Qr \sin(Qr)\, dQ$$

The transformation was implemented using GNOM software (Semenyuk & Svergun, 1991). For the C3-FH and C3u-FH mixtures in 137 mM and 50 mM NaCl, the X-ray $I(Q)$ curves with 242-394 data points used the Q range between 0.10 nm^{-1} and extending up to 1.61-1.85 nm^{-1}.

(6.4.4) Constrained scattering and sedimentation coefficient modelling

The constrained modelling of the experimental scattering curves for the C3-FH and C3u-FH mixtures required molecular models for each component present in the mixture in order to initiate this procedure. The modelling of the non-interacting C3-FH mixtures in 137 mM NaCl and 50 mM NaCl was based on two components. These were the crystal structure for human C3 (PDB code 2A73, Jenssen *et al.,* 2005) and the solution structure for FH (PDB code 3N8O, Okemefuna *et al.,* 2010a). The modelling of the interaction between C3u and FH in 137 mM NaCl was based on three components. These were the solution structures of human C3u (PDB code 3MMQ, Li *et al.,* 2010; Chapter 5) and FH (PDB code 3N8O, Okemefuna *et al.,* 2010a), together with the crystal structure of C3b in complex with FH SCR-1/4 (PDB code 2WII, Wu *et al.*, 2009). That FH solution structure was the most-recently determined one based on the constrained modelling searches of the SCR-1/5, SCR-6/8, SCR-8/11, SCR-11/15 and SCR-16/20 fragments whose structures were previously determined (PDB codes 2QFG, 2QFH, 2UWN, 3N8P, 3N8Q, Okemefuna *et al.,* 2010a). The solution structure of the 1:1 C3u-FH complex was approximated by the C3b-FH complex. This structure was formed by the superimposition of SCR-4 in the C3b-FH-SCR-1/4 crystal structure upon the SCR-4 domain of the solution structure of FH, after which the SCR-1/4 domains in

the FH solution structure model were deleted. The modelling of the interaction between C3u and FH in 50 mM NaCl was based on five components. Three were taken from the solution structural models of C3u and FH and their 1:1 C3u-FH complex (above). The fourth was represented by the solution structure of the $C3u_2$-FH complex, which was generated starting from the crystal structure of the C3b dimer (PDB code 2WIN, (Rooijakkers *et al.,* 2009). The MG-1/8 domains of one monomer in this C3b dimer were superimposed upon the MG-1/8 domains in the 1:1 C3u-FH model, after which the C3b monomer within that dimer was removed to result in a 2:1 C3u-$_2$FH structural model. The fifth was represented by a model for the C3u-FH-C3u complex, in which C3u interacted at two independent FH sites at SCR-1/4 and SCR-19/20. For this, a C3u structure was positioned close to the SCR-19/20 domains in the 1:1 C3u-FH model (above) was adapted. This positioning was guided according to binding sites identified from previous mutation studies and sequence analyses (Table 6.2). Thus the residues K1187- A1249 in the TED domain of C3b were placed close to the residues R1182, W1183, R1210 and R1215 of SCR-20 and W1157 of SCR 19.

Each of the five molecular models was used to calculate its X-ray scattering curve following their transformation into Debye sphere models (Perkins & Weiss, 1983). For C3, C3u and FH, cube side lengths of 0.540 nm, 0.540 nm and 0.498 nm in conjuction with a cut off of 4 atoms were used. For C3u-FH, $C3u_2$-FH and C3u-FH-C3u, a cube side length of 0.762 nm in conjunction with a cutoff of 9 atoms was used. Since these structures are unhydrated, and correspond to the amino acid and carbohydrate residues expected to be present, hydration in each sphere model was achieved by cover the surface of the sphere model with additional spheres to correspond to 0.3 g H_2O/ g glycoprotein using HYPRO (Ashton *et al*., 1997). The optimum totals of dry and hydrated spheres were 1557 and 2047 (C3u), 1564 and 2074 (FH), 991 and 1308 (C3u-FH) and 1547 and 2038 ($C3u_2$-FH and C3u-FH-C3u) respectively. The X-ray scattering curve *I(Q)* was calculated using the Debye equation adapted to spheres (Boehm *et al*., 1999; Perkins, 2001b; Li *et al*., 2010). The X-ray curves were calculated without instrumental corrections because these are considered to be negligible for the pinhole optics used in synchrotron X-ray scattering instruments. In order to model the experimental X-ray scattering curve for C3-FH mixtures in 137 mM and 50 mM NaCl, for which C3 and FH do not interact, the C3 and FH scattering curves were merged in increments of 1% from 0% to 100% of C3 and FH in order to generate 101 predicted curves for comparison with the experimental X-ray curves for each of the two buffers,

following the method used to analyse pentamer-decamer mixtures of C-reactive protein (Okemefuna *et al.,* 2010b). In order to model the experimental X-ray scattering curve for C3u-FH mixtures in 137 mM NaCl, where a 1:1 complex was formed, the three individual curves for C3u, FH and the 1:1 complex were merged in 2% incremental steps to generate 1326 predicted curves for comparison with three different experimental X-ray curves in 137 mM NaCl. In order to model the experimental X-ray scattering curve for C3u-FH mixtures in 50 mM NaCl, where a 1:1 complex and two 2:1 complexes were formed, the five individual curves for C3u, FH, the 1:1 complex and the two 2:1 complexes were merged in 2% incremental steps from 0% to 100% to generate 316,260 predicted five-component curves for comparison with five different experimental X-ray curves in 50 mM NaCl. The three sets of predicted curves for the four experiments were evaluated to identify the proportions that resulted in the best fit to the experimental curves. Firstly, the R_G values from the predicted curves were required to be within 5% of the experimental values, using Guinier fits in the same Q ranges used for the experimental Guinier analyses. Secondly, the overall goodness-of-fit R-factor between the predicted and experimental curves were calculated in order to identify the best-fit model from the lowest R-factor.

Sedimentation coefficients $s^0_{20,w}$ were calculated directly from the hydrated Debye sphere models using the program HYDRO (Garcia de la Torre *et al.,* 1994). They were also calculated from the atomic coordinates for each model using the HYDROPRO shell modelling program using the default value of 0.31 nm for the atomic element radius for all atoms to represent the hydration shell (Garcia de la Torre *et al.,* 2000). Previous applications of these calculations to proteins

Chapter Seven

Conclusions

7.1 Prologue

The projects presented in this thesis have defined the solution structures and associations of CR2 SCR 1-15, C3d, C3 and C3u, and have mainly focused on the weak interactions between CR2-C3d and C3u-FH. These weak interactions were researched both in low salt buffers, where a stronger interaction is promoted, and in a more physiological salt buffer. The major projects in this thesis investigated the structure of CR2 SCR 1-15 and the effect of C3d on this, and also the structure of C3u and the interaction between C3u and FH and placed the latter in the context of the relationship to C3b and the C3b-FH complex.

Protein-protein interactions can be strong with 100% of complex formed (with nanomolar dissociation constants) or weak ones where the complex is partially dissociated (with micromolar dissociation constants) (Perkins *et al.*, 2010a). The majority of the complement proteins function through their relative abundance and their weak interactions. The advantage of AUC is that in principle this identifies all the species in solution through *c(s)* size-distribution analyses; however, interpretations need to be cautious in case reaction boundaries are present (Cole *et al.*, 2008). AUC is a powerful method for the protein interaction studies in this thesis because AUC extracts K_D values from sedimentation equilibrium and sedimentation velocity fits, and displays all the components present from $c(s)$ distribution analyses. X-ray scattering is a technique which is most powerful when it is used to extract medium-resolution structural information through constrained modelling on the basis of atomic resolution structures. Although the scattering technique works best if the sample is monodisperse (Perkins et al., 2008, 2009), the combination of AUC and scattering methods takes advantage of the strengths of both approaches. Interestingly, it proved possible in this thesis to use scattering modelling (Chapter 6) with the combination of the scattering curves from the individual components to give independent information on the stoichiometry and the K_D of a complex protein interaction.

The solution structures of CR2 and C3u suggested that flexibility is important for these two large complement proteins (Chapter 4, Chapter 5). For FH, there are reasons to believe that flexibility may be a less significant property. To evaluate this, the combination of atomic resolution structures and solution data complement each other, and this combination was most useful for understanding the relative contributions of structure and flexibility to function in these proteins.

7.2 Interaction of CR2 SCR 1-15 with C3d

The knowledge of how CR2 bound with C3d from this thesis provided clearer insight into the function of CR2. Here, the effect of C3d binding to full-length CR2 was determined by a joint AUC and X-ray scattering study (Chapter 4). First, it was necessary to understand the behaviour of the free proteins. The sedimentation coefficient of unbound CR2 (4.03 S) in 50 mM NaCl agrees well with that in 137 mM NaCl (3.93 S), so the overall CR2 structure is unaffected by change in ionic strength (Section 4.2.1). Previously it was presumed that unbound C3d underwent dimerisation in 50 mM NaCl buffer (Gilbert *et al*., 2005). In this thesis,my study of sedimentation velocity and sedimentation equilibrium data illustrated that this self-association is best described by two monomer-dimer and monomer-trimer equilibria in 50 mM NaCl, but as a monomer only in 137 mM NaCl instead (Figure 4.3 and 4.4).

For the CR2-C3d complex, the *c*(*s*) size-distribution analyses by AUC using 50 mM NaCl showed that the complex is formed (Figure 4.2); while in 137 mM NaCl the *c*(*s*) analyses revealed dissociated components (Section 4.2.4). The previous solution structural models for the CR2 SCR-1/2 complex with C3d (Gilbert *et al*., 2005) was superimposed on the solution structure for full-length CR2 SCR-1/15. This model gave an average sedimentation coefficient of 4.57 S for the complex, in good agreement with the observed value of 4.52 S (Figure 4.6; Table 4.1). It is concluded that CR2 did not detectably change conformation when C3d was bound to it, and also that the interaction of CR2 with C3d demonstrates a substantial dependence on charge. This outcome agreed with surface plasmon resonance studies in 50 mM NaCl that included SCR 1-2, SCR 1-4 and SCR 1-15 which demonstrated a high affinity for C3d, but in more physiological 125 mM NaCl salt, the binding interaction between CR2 and C3d was much weakened (Asokan *et al.,* 2006; Moore *et al.,* 1989; Guthridge *et al.,* 2001; Hannan *et al.,* 2005). To our knowledge, this is the first solution structural study of a large multidomain SCR protein CR2 bound to its physiological ligand C3d. Consistent with previous analyses, the CR2-C3d complex is not formed in physiological salt conditions. The lack of an observed interaction by AUC that is known to exist *in vivo* is best explained by the physiological context of this interaction. CR2 is a membrane bound protein, and will interact with an abundance of C3d fragments that are also cell surface bound. Such an interaction between two surfaces would be promoted by local surface concentration effects.

7.3 Solution structures of C3 and C3u

The same strategy was applied to study the C3u-FH interaction. First, the self-association properties and solution structures of each of C3 and C3u were determined using AUC and X-ray scattering (Chapter 5). Sedimentation coefficients identified two different types of dimerisation events in both proteins. A fast dimerisation was observed in 50 mM NaCl that was removed in 137 mM NaCl (Figure 5.3(e,f)). A low amount of a slow dimerisation was observed for C3u and C3 in both buffers (Figure 5.5). The X-ray radius of gyration R_G values were unchanged for both C3 and C3u in 137 mM NaCl, but depend on concentrations in 50 mM NaCl (Figure 5.7). Solution structures were also determined. The C3 crystal structure gave good X-ray scattering fits for C3 in 137 mM NaCl. For C3u, constrained modelling in 137 mM NaCl showed that the TED/CUB domains in C3u were extended and differ from the more compact arrangement seen in the crystal structures of C3b that were crystallised in low salt (Figure 5.15 (a,b)). This TED/CUB conformation is intermediate between those seen in the C3 and C3b crystal structures. The greater exposure of the TED domain in C3u (which possesses the hydrolysed reactive thioester) accounts for the greater self-association of C3u in low salt. In addition, the conformational variability of the functionally-important TED/CUB domains would facilitate their interactions with a broad range of antigenic surfaces. One of the two dimerisation events seen for C3 and C3u may correspond to a dimer observed in one of the crystal structures of C3b.

7.4 Interactions of C3/C3u with Factor H in solution

To study the C3u-FH interaction itself, the C3u-FH interaction (Chapter 6) was studied by sedimentation velocity experiments that showed that both 1:1 and 2:1 complexes of C3u-FH were observed in both buffers. In 137 mM NaCl buffer, the 1:1 complex is predominant, while the relative proportion of 2:1 complexes was increased in 50 mM NaCl buffer (Figure 6.3(d,h)). Dissociation constants K_D of 0.59 μM and 0.11 μM were determined in these buffers from integrations of the $c(s)$ plots. Starting from previously-determined structures for FH, C3u, the FH (SCR-1/4)-C3b complex and the C3b dimer, a novel scattering curve fit procedure was developed for weakly-interacting macromolecules. For 137 mM NaCl, a three-component system gave the best curve fit, with 48% of unbound C3u and FH, and 52% 1:1 complex (Table 6.3), leading to a similar K_D value of 0.22 μM (Table 6.4). In 50 mM NaCl, good fits were obtained using a five-component analysis with a 1:1 complex and two different types of 2:1 complexes (Figure 6.11). In summary, the ability to model the sedimentation coefficient of the 1:1

complex showed that no conformational changes in FH were detected in the C3u-FH complexes (Figure 6.9). In addition, two independent binding sites for C3u were shown to exist in FH, even though it is often assumed in the literature that one FH molecule is bent around one molecule of C3u when the FH-C3u complex is formed. Our results provide new insight on how FH achieves regulatory control of C3u activation.

7.5 Biological significance

Complement activation causes inflammation in the host, so control mechanisms are crucial for protection of the host cells. This includes ways to stop the formation of excess amounts of the C3 convertase by promoting the decay of activated C3b. My thesis illustrates newer insights on how FH and CR2 play a role in the regulation of C3 and its fragments. The combination of AUC, X-ray scattering analyses and constrained modelling has established the solution structures of CR2-C3d, C3u and C3u-FH. These findings are related to many broader aspects of complement function. Thus, in the alternative pathway, FH regulates the active central complement protein C3b. C3b shares functional similarity with C3u (also known as $C3_{H20}$), thus the studies of the C3u-FH interactions are expected to be useful in terms of the C3b-FH complex. The results on the flexibility of the TED-CUB domains in C3b are expected to apply to the TED-CUB domains in the related proteins C4b and C5b.

An understanding of the complement protein interactions will clarify the role of complement in disease. For example, it was suggested that the complement-mediated antibody-dependent enhancement of HIV infection is mediated by the association of CR2 expressed on FDC or B cells with the C3d fragment attached to the surface of HIV virons (Xu *et al.*, 2009). Depletion of FH from serum dramatically increases the sensitivity of HIV to complement-mediated lysis, thus enables the HIV virions to spread more rapidly; however in the presence of FH, C3b is regulated and lysis will be inhibited (Stoiber *et al.*, 1996, 2001; Willey & Assa-Chapman, 2008). The studies of the CR2-C3d and C3u-FH interactions in this thesis build a clearer foundation by which the HIV infection and blockage mechanism operates. Additionally, the solution structure of the C3u-FH complexes provides new insight on the involvement of FH in atypical haemolytic uraemic syndrome (aHUS), where mutations in FH are associated with aHUS. In summary, the solution structures of this thesis lead to a greater understanding of complement activation regulation and possibilities for new drug discoveries.

References

Aasa-Chapman, M. M., Holuigue, S., Aubin, K., Wong, M., Jones, N. A., Cornforth, D., Pellegrino, P., Newton, P., Williams, I., Borrow, P. & McKnight, A. (2005). Detection of antibody-dependent complement-mediated inactivation of both autologous and heterologous virus in primary human immunodeficiency virus type 1 infection. *J. Virol.* **79**, 2823-2830.

Ahearn, J. M., Hayward, S. D., Hickey, J. C. & Fearon, D. T. (1988). Epstein-Barr virus (EBV) infection of murine L cells expressing recombinant human EBV/C3d receptor. *Proc. Natl. Acad. Sci. U. S. A.* **85**, 9307-9311.

Ahearn, J. M., Fischer, M. B., Croix, D., Goerg, S., Ma, M., Xia, J., Zhou, X., Howard, R. G., Rothstein, T. L. & Carroll, M. C. (1996). Disruption of the c*r2* locus results in a reduction in B-1a cells and in an impaired B cell response to T-dependent antigen. *Immunity*. **4**, 251-262.

Ajees, A. A., Gunasekaran, K., Volanakis, J. E., Narayana, S. V. L., Kotwal, G. J. & Murthy, K. H. M. (2006). The structure of complement C3b provides insights into complement activation and regulation. *Nature,* **444**, 221-225.

Ajees, A. A., Gunasekaran, K., Narayana, S. V. L. & Murthy, K. H. M. (2007). Reply. *Nature*, **448**, E2-E3.

Alper, C. A., Johnson, A. M., Birtch, A. G. & Moore, F. D. (1969). Human C3: evidence for the liver as the primary site of synthesis. *Science*, **163**, 286-288.

Alsenz, J., Lambris, J. D., Schulz, T. F. & Dierich, M. P. (1984). Localization of the complement-component-C3b-binding site and the cofactor activity for factor I in the 38kDa tryptic fragment of factor H. *Biochem. J.* **224**, 389–398.

Alsenz, J., Schulz, T. F., Lambris, J. D., Sim, R. B. & Dierich, M. P. (1985). Structural and functional analysis of the complement component factor H with the use of different enzymes and monoclonal antibodies to factor H. *Biochem. J.* **232**, 841-50.

Armstrong, P. B., Armstrong, M. T. & Quigley, J. P. (1993) Involvement of alpha 2-macroglobulin and C-reactive protein in a complement-like hemolytic system in the arthropod, Limulus polyphemus. *Mol. Immunol.* **30**, 929-934.

Ashton, A. W., Boehm, M. K., Gallimore, J. R., Pepys, M. B. & Perkins, S. J. (1997). Pentameric and decameric structures in solution of the serum amyloid P component

by X-ray and neutron scattering and molecular modelling analyses. *J. Mol. Biol*. **272**, 408-422.

Aslam, M. & Perkins, S. J. (2001). Folded-back solution structure of monomeric factor H of human complement by synchrotron X-ray and neutron scattering, analytical ultracentrifugation and constrained molecular modelling. *J. Mol. Biol.* **309**, 1117-1138.

Aslam, M., Guthridge, J. M., Hack, B. K., Quigg, R. J., Holers, V. M. & Perkins, S. J. (2003). The extended multidomain solution structures of the complement protein Crry and its chimaeric conjugate Crry-Ig by scattering, analytical ultracentrifugation and constrained modelling: implications for function and therapy. *J. Mol. Biol*. **329**, 525-550.

Asokan, R., Hua, J., Young, K. A., Gould, H. J., Hannan, J. P., Kraus, D. M., Szakonyi, G., Grundy, G. J., Chen, X. S., Crow, M. K. & Holers, V. M. (2006). Characterization of human complement receptor type 2 (CR2/CD21) as a receptor for IFN-alpha: a potential role in systemic lupus erythematosus. *J. Immunol*. **177**, 383-394.

Aubry, J. P., Pochon, S., Graber, P., Jansen, K. U. & Bonnefoy, J. Y. (1992). CD21 is a ligand for CD23 and regulates IgE production. *Nature*, **358**, 505-507.

Aubry, J. P., Pochon, S., Gauchat, J. F., Nueda-Marin, A., Holers, V. M., Graber, P., Siegfried, C. & Bonnefoy, J. Y. (1994). CD23 interacts with a new functional extracytoplasmic domain involving N-linked oligosaccharides on CD21. *J. Immunol*. **152**, 5806-5813.

Balbo, A. & Schuck, P. (2005). Analytical ultracentrifugation in the study of protein self-association and heterogeneous protein-protein interactions. In *Protein-Protein Interactions* (Golemis, E. & Adams, P. D., eds), pp. 253-277. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, NY.

Bartók, I. & Walport, M. J. (1995). Comparison of the binding of C3S and C3F to complement receptors types 1, 2, and 3. *J. Immunol*. **154**, 5367-5375.

Basset-Séguin, N., Caughman, S. W. & Yancey, K. B. (1990). A-431 cells and human keratinocytes synthesize and secrete the third component of complement. *J. Invest. Dermatol*. **95**, 621-625.

Beckman Model XL-A/XL-I Analytical Ultracentrifuge Optima^{TM} series Training Guide (1998). Beckman Coulter, High Wycombe, United Kingdom.

Bhattacharjee, A., Lehtinen, M. J., Kajander, T., Goldman, A. & Jokiranta, T. S. (2010). Both domain 19 and domain 20 of factor H are involved in binding to complement C3b and C3d. *Mol. Immunol.* **47**, 1686-1691.

Birkenbach, M., Tong, X., Bradbury, L. E., Tedder, T. F. & Kieff, E. (1992). Characterization of an Epstein-Barr virus receptor on human epithelial cells. *J. Exp. Med.* **176**, 1405-1414.

Bitter-Suermann, D. & Burger, R. (1990). C3 deficiencies. *Curr. Top. Microbiol. Immunol*. **153**, 223-233.

Blackmore, T. K., Sadlon, T. A., Ward, H. M., Lublin, D. M. & Gordon, D. L. (1996). Identification of a heparin binding domain in the seventh short consensus repeat of complement factor H. *J. Immunol*. **157**, 5422-5427.

Blackmore, T. K., Hellwage, J., Sadlon, T. A., Higgs, N., Zipfel, P. F., Ward, H. M. & Gordon, D. L. (1998). Identification of the second heparin binding domain in human complement factor H. *J. Immunol*., **160**, 3342-3348.

Blandin, S. & Levashina, E. A. (2004). Thioester-containing proteins and insect immunity, *Mol. Immunol.* **40**, 903-908.

Boehm, M. K., Woof, J. M., Kerr, M. A. & Perkins, S. J. (1999). The Fab and Fc fragments of IgA1 exhibit a different arrangement from that in IgG: a study by X-ray and neutron solution scattering and homology modelling. *J. Mol. Biol.* **286**, 1421-1447.

Bokisch, V. A., Müller-Eberhard, H. J. & Cochrane, C. G. (1969). Isolation of a fragment (C3a) of the third component of human complement containing anaphylatoxin and chemotactic activity and description of an anaphylatoxin inactivator of human serum. *J. Exp. Med*. **129**, 1109-1130.

Bonner, A., Almogren, A., Furtado, P. B., Kerr, M. A. & Perkins, S. J. (2009). Location of secretory component on the Fc edge of dimeric IgA1 reveals insight into the role of secretory IgA1 in mucosal immunity. *Mucosal. Immunol.* **2**, 74-84.

Bordet, J. & Gengou. O. (1901). Sur l'existences de substance sensibilisations. *Ann. Inst. Past.* **15**, 290.

Bordet, J. (1909). In *Studies in Immunity*. J. Wiley & Sons, New York.

Borrell, B. (2009). Fraud rocks protein community. *Nature*, **462**, 970-970.

Botto, M., Fong, K. Y., So, A. K., Rudge, A. & Walport, M. J. (1990). Molecular basis of hereditary C3 deficiency. *J. Clin. Invest*. **86**, 1158-1163.

Botto, M., Fong, K. Y., So, A. K., Barlow, R., Routier, R., Morley, B. J. & Walport, M. J. (1992). Homozygous hereditary C3 deficiency due to a partial gene deletion. *Proc. Natl. Acad. Sci. U. S. A*. **89**, 4957-4961.

Bradbury, L. E., Kansas, G. S., Levy, S., Evans, R. L. & Tedder, T. F. (1992). The CD19/CD21 signal transducing complex of human B lymphocytes includes the target of antiproliferative antiproliferative antibody-1 and Leu-13 molecules. *J. Immuno*l. **149**, 2841-2850.

Brown, P. H. & Schuck P. (2006). Macromolecular size-and-shape distributions by sedimentation velocity analytical ultracentrifugation. *Biophys. J.* **90**, 4651-4661.

Brown, K. M., Kondeatis, E., Vaughan, R. W., Kon, S. P., Farmer, C. K., Taylor, J. D., He, X., Johnston, A., Horsfield, C., Janssen, B. J., Gros, P., Zhou, W., Sacks, S. H. & Sheerin, N. S. (2006). Influence of donor C3 allotype on late renal-transplantation outcome. *N. Engl. J. Med*. **354**, 2014-2023.

Buchner, H. (1889). Über die nähere Natur der bakterientötenden Substanz in Blutserum. *Zbl. Bakt.* **6**, 561-565.

Budd, A., Blandin, S., Levashina, E. A. & Gibson, T. J. (2004). Bacterial alpha2-macroglobulins: colonization factors acquired by horizontal gene transfer from the metazoan genome? *Genome. Biol.* **5**, R38.

Cantor, C. R. & Schimmel, P. R. (1980). Biophysical Chemistry, part II: Techniques for the Study of Biological Structure and Function. W. H. Freeman & Company ed., San Francisco.

Carel, J. C., Myones, B. L., Frazier, B. & Holers, V. M. (1990). Structural requirements for C3dg/Epstein-Barr Virus Receptor (CR2/CD21) ligand binding, internalization and viral infection. *J. Biol. Chem*. **265**, 12293-12299.

Carroll, M. C. (2004). The complement system in B cell regulation. *Mol. Immunol*. **41**, 141-6. Review.

Carter, R. H., Spycher, M. O., Hg, Y. C., Hoffman, R. & Fearon, D. T. (1988). Synergistic interaction between complement receptor type 2 and membrane IgM on B lymphocytes. *J. Immunol*. **141**, 457-463.

Carter, R. H., Tuveson, D. A., Park, D. J., Rhee, S. G. & Fearon, D. T. (1991). The CD19 complex of B lymphocytes. Activation of phospholipase C by a protein tyrosine kinase-dependent pathway that can be enhanced by the membrane IgM complex. *J. Immunol.* **147**, 3663-3671.

Carter, R. H. & Fearon, D. T. (1992). CD19: Lowering the threshold for antigen receptor stimulation of B lymphocytes. *Science,* **256**, 105-107.

Chacon, P., Moran, F., Diaz, J. F., Pantos, E. & Andreu J. M. (1998). Low-resolution structures of proteins in solution retrieved from X-ray scattering with a genetic algorithm. *Biophys. J*. **74**, 2760-2775.

Chang, V. T., Crispin, M., Aricescu, A. R., Harvey, D. J., Nettleship, J. E., Fennelly, J. A., Yu, C., Boles, K. S., Evans, E. J., Stuart, D. I., Dwek, R. A., Jones, E. Y., Owens, R. J. & Davis, S. J. (2007) Glycoprotein structural genomics: solving the glycosylation problem. *Structure,* **15**, 267-273.

Chu, C. T. & Pizzo, S. V. (1994). Alpha 2- Macroglobulin, complement, and biologic defense: antigens, growth factors, microbial proteases, and receptor ligation. *Lab. Invest*. **71**, 792-812.

Clemenza, L. & Isenman, D. E. (2000). Structure-guided identification of C3d residues essential for its binding to complement receptor 2 (CD21). *J. Immunol*. **165**, 3839-3848.

Coca, A. F. (1914). A study of the anticomplementary action of yeast on certain bacteria and of cobra venom. *Z. Immunitätsforsch.* **21**, 604-610.

Cole, J. L. & Hansen, J. C. (1999). Analytical ultracentrifugation as a contemporary biomolecular research tool. *J. Biomol. Tech.* **10**, 163-176.

Cole, J. L. (2004). Analysis of heterogeneous interactions. *Methods. Enzymol*. **384**, 212-232.

Cole, J. L., Lary, J. W., Moody, T. P. & Laue, T. M. (2008). Analytical ultracentrifugation: sedimentation velocity and sedimentation equilibrium. *Meth. Cell. Biol.* **84**, 143-211.

Cölfen, H. & Völkel, A. (2003). Hybrid colloid analysis combining analytical ultracentrifugation and flow-field flow fractionation. *Eur. Biophys. J*. **32**, 432-436.

Correia, J. J. (2000). Analysis of weight average sedimentation velocity data. *Methods Enzymol*. **321**, 81-100.

Correia, J. J., Sontag, C. A., Stafford, W. F. & Sherwood, P. J. (2005). Models for Direct Boundary Fitting of Indefinite Ligand-Linked Self-Association, In *Analytical Ultracentrifugation: Techniques and Methods*. (Scott, D. J., Harding, S. E. & Rowe, A. J., eds), pp. 51-63. Royal Society of Chemistry, Cambridge.

Creeth, J. M. & Knight, C. G. (1965). On the estimation of the shape of macromolecules from sedimentation and viscosity measurements. *Biochim Biophys Acta*. **102**, 549-558.

Dam, J. & Schuck, P. (2004). Calculating sedimentation coefficient distribution by direct modeling of sedimentation velocity concentration profiles. *Methods Enzymol*. **384**, 185-212.

Dam, J. & Schuck, P. (2005). Sedimentation velocity analysis of heterogeneous protein-protein interactions: sedimentation coefficient distributions *c(s)* and asymptotic boundary profiles from Gilbert-Jenkins theory. *Biophys. J.* **89**, 651-666.

Dam, J., Velikovsky, C. A., Mariuzza, R. A., Urbanke, C. & Schuck, P. (2005). Sedimentation velocity analysis of heterogeneous protein-protein interactions: Lamm equation modeling and sedimentation coefficient distributions *c*(*s*). *Biophys. J.* **89**, 619-634.

Davis III, A. E. (1998). C1 Inhibitor Gene and Hereditary Angioedema. In *The Human Complement System in Health and Disease*. (Volanakis, J. E. & Frank, M. M., eds), pp. 455-480. Marcel Dekker Inc., New York.

De Bruijn, M. H. & Fey, G. H. (1985).Human complement component C3: cDNA coding sequence and derived primary structure. *Proc. Natl. Acad. Sci. U. S. A*. **82**, 708-712.

Debye, P. (1915). Zerstreuung von röntgenstrahlen. Scattering from noncrystalline substances. *Ann. Phys*. **46**, 809–823.

Delcayre, A. X., Salas, F., Mathur, S., Kovats, K., Lotz, M. & Lernhardt, W. (1991). Epstein Barr virus/complement C3d receptor is an interferon alpha receptor**.** *EMBO. J.* **10**, 919-926.

Dempsey, P. W., Allison, M. E. D., Akkaraju, S., Goodnow, C. C. & Fearon, D. T. (1996). C3d of complement as a molecular adjuvant: Bridging innate and acquired immunity. *Science*, **271**, 348-350.

Dodds, A, W. (1993). Small-scale preparation of complement components C3 and C4. *Methods Enzymol.* **223**, 46-61.

Drenth, J. (1999) Principle of protein x-ray crystallography. Springer-Verlag, New York. pp. 301.

Ellman, L., Green, I. & Frank, M. M. (1970). Genetically controlled total deficiency of the fourth component of complement in the guinea pig. *Science,* **170**, 74-75.

Epstein, M. A., Achong, B. G. & Barr, Y. M. (1964). Virus particles in cultured lymphoblasts from Burkitt's lymphoma. *Lancet*, **1**, 702-703.

Fearon, D. T. & Carroll, M. C. (2000). Regulation of B lymphocyte responses to foreign and self-antigens by the CD19/CD21 complex. *Annu. Rev. Immunol*. **18**, 393-422.

Fenaille, F., Le Mignon, M., Groseil, C., Ramon, C., Riandé, S., Siret, L. & Bihoreau, N. (2007). Site-specific N-glycan characterization of human complement factor H. *Glycobiol*. **17**, 932-944.

Feramisco, J. R., Glass, D. B. & Krebs, E. G. (1980). Optimal spatial requirements for the location of basic residues in peptide substrates for the cyclic AMP-dependent protein kinase. *J. Biol. Chem.* **255**, 4240-4245.

Fernando, A. N., Furtado, P. B., Clark, S. J., Gilbert, H. E., Day, A. J., Sim, R. B., & Perkins, S. J. (2007). Associative and structural properties of the region of complement Factor H encompassing the Tyr402His disease-related polymorphism and its interactions with heparin. *J. Mol. Biol.* **368**, 564-581.

Ferrata, A. (1907). Die unwirksamkeit der komplexen hämolysine in salzfreien lösungen und ihre ursache, *Berlin, Klin. Wochenschr*. **44**, 366.

Ferreira, V. P., Herbert, A. P., Hocking, H. G., Barlow, P. N. & Pangburn, M. K. (2006). Critical role of the C-terminal domains of factor H in regulating complement activation at cell surfaces. *J. Immunol.* **177**, 6308-6316.

Fingeroth, J. D., Weis, J. J., Tedder, T. F., Strominger, J. L., Biro, P. A. & Fearon, D. T. (1984). Epstein-Barr virus receptor of human B lymphocytes is the C3d receptor CR2. *Proc. Natl. Acad. Sci. U. S. A.* **81**, 4510-4514.

Fischer, E., Delibrias, C. & Kazatchkine, M. D. (1991). Expression of CR2 (the C3dg/EBV receptor, CD21) on normal human peripheral blood T lymphocytes. *J. Immunol.* **146**, 865-869.

Fredslund, F., Jenner, L., Husted, L. B., Nyborg, J., Andersen, G. R. & Sottrup-Jensen, L. (2006). The structure of bovine complement component 3 reveals the basis for thioester function. *J. Mol. Biol.* **361**, 115-127.

Fredslund, F., Laursen, N. S., Roversi, P., Jenner, L., Oliveira, C. L., Pedersen, J. S., Nunn, M. A., Lea, S. M., Discipio, R., Sottrup-Jensen, L. & Andersen, G. R. (2008). Structure of and influence of a tick complement inhibitor on human complement component 5. *Nat. Immunol.* **9**, 753-760.

Frémeaux-Bacchi, V., Bernard, I., Maillet, F., Mani, J. C., Fontaine, M., Bonnefoy, J. Y., Kazatchkine, M. D. & Fischer, E. (1996). Human lymphocytes shed a soluble form of CD21 (the C3dg/Epstein-Barr virus receptor, CR2) that binds iC3b and CD23. *Eur. J. Immunol.* **26**, 1497-1503.

Frémeaux-Bacchi, V., Miller, E. C., Liszewski, M. K., Strain, L., Blouin, J., Brown, A. L., Moghal, N., Kaplan, B. S., Weiss, R. A., Lhotta, K., Kapur, G., Mattoo, T., Nivet, H., Wong, W., Gie, S., Hurault de Ligny, B., Fischbach, M., Gupta, R., Hauhart, R., Meunier, V., Loirat, C., Dragon-Durey, M. A., Fridman, W. H., Janssen, B. J., Goodship, T. H. & Atkinson, J. P. (2008). Mutations in complement C3 predispose to development of atypical hemolytic uremic syndrome. *Blood,* **112**, 4948-4952.

Fujisaku, A., Harley, J. B., Frank, M. B., Gruner, B. A., Frazier, B. & Holers, V. M. (1989). Genomic organization and polymorphisms of the human C3d/Epstein-Barr virus receptor. *J. Biol. Chem.* **264**, 2118-2125.

Furtado, P. B., Huang, C. Y., Ihyembe, D., Hammond, R. A., Marsh, H. C. & Perkins, S. J. (2008). The partly-folded back solution structure arrangement of the 30 SCR domains in human complement receptor type 1 (CR1) permits access to its C3b and C4b ligands. *J. Mol. Biol*. **375**, 102-118.

Ganu, V. S., Müller-Eberhard, H. J. & Hugli, T. E. (1989). Factor C3f is a spasmogenic fragment released from C3b by factors I and H: the heptadeca-peptide C3f was synthesized and characterized. *Mol. Immunol.* **26**, 939-948.

Garcia de la Torre, J., Navarro, S., Martinez, M. C. L., Diaz, F. G. & Cascales, J. L. (1994). HYDRO: A computer program for the prediction of hydrodynamic properties of macromolecules. *Biophys. J*. **67**, 530-531.

Garcia de la Torre, J. & Bloomfield, V. A. (1997). Hydrodynamic properties of macromolecular complexes. I. Translation. *Biopolymers,* **16**, 1747-1761.

Garcia de la Torre, J., Huertas, M. L. & Carrasco, B. (2000). Calculation of hydrodynamic properties of globular proteins from their atomic-level structure. *Biophys. J*. **78**, 719-730.

Gasque, P. Chan. P., Mauger, C., Schouft, M. T., Singhrao, S., Dierich, M. P., Morgan, B. P. & Fontaine M. (1996). Identification and characterization of complement C3 receptors on human astrocytes. *J. Immunol.* **156**, 2247-2255.

Giebeler, R. (1992). The optima XL-A: a new analytical ultracentrifuge with a novel precision absorption optical system. In *Analytical Ultracentrifugation in Biochemistry and Polymer Science* (Harding, S. E., Rowe, A. J. & Horton, J. C. eds). pp.16–25. Royal Society of Chemistry, Cambridge.

Gilbert, H. E., Eaton, J. T., Hannan, J. P., Holers, V. M. & Perkins, S. J. (2005). Solution structure of the complex between CR2 SCR 1-2 and C3d of human complement: an X-ray scattering and sedimentation modelling study. *J. Mol. Biol.* **346**, 859-873.

Gilbert, H. E., Aslam, M., Guthridge, J. M., Holers, V. M. &. Perkins, S. J. (2006a). Extended linkers between CR2 SCR-1 and SCR-2 and the Fc fragment in CR2-Ig by X-ray and neutron scattering, analytical ultracentrifugation and constrained molecular modelling. *J. Mol. Biol.* **356**, 397-412.

Gilbert, H. E., Asokan, R., Holers, V. M. & Perkins, S. J. (2006b). The 15 SCR flexible extracellular domains of human complement receptor type 2 can mediate multiple ligand and antigen interactions. *J. Mol. Biol.* **362**, 1132-1147.

Gilbert, H. E. (2006). Structural studies of SCR domains in complement receptor type two. Ph.D thesis. University College London.

Glatter, O. & Kratky, O. (1982). Editors of *Small-angle X-ray scattering*. Academic Press, New York.

Gordon, J., Whitehead, H. R. & Wormall, A. (1926). The action of ammonia on complement. The fourth component. *Biochem. J*. **20**, 1028-35.

Gordon, D. L., Kaufman, R. M., Blackmore, T. K., Kwong, J. & Lublin, D. M. (1995). Identification of complement regulatory domains in human factor H. *J. Immunol*. **155**, 348-356.

Götze, O. & Müller-Eberhard, H. J. (1971). The C3-activator system: An alternative pathway of complement activation. *J. Exp. Med.* **134**, 90-108.

Green, T. D., Montefiori, D. C. & Ross, T. M. (2003). Enhancement of antibodies to the human immunodeficiency virus type 1 envelope by using the molecular adjuvant C3d, *J. Virol*. **77**, 2046-2055.

Guinier, A. & Fournet, G. (1955). Small-angle scattering of X-rays. Wiley, New York.

Guthridge, J. M., Rakstang, J. K., Young, K. A., Hinshelwood, J., Aslam, M., Robertson, A., Gipson, M. G., Sarrias, M.-R., Moore, W. T., Meagher, M., Karp, D., Lambris, J. D., Perkins, S. J. & Holers, V. M. (2001). Structural studies in solution of the recombinant N-terminal pair of short consensus/complement repeat domains of complement receptor type 2 (CR2/CD21) and its interaction with its ligand C3dg. *Biochemistry,* **40**, 5931-5941.

Halle, B. (2004).Protein hydration dynamics in solution: a critical survey. *Philos. Trans. R. Soc. Lond. B*. **359**, 1207-1224

Hammel, M., Kriechbaum, M., Gries, A., Kostner, G. M., Laggner P. & Prassl, R. (2002). Solution structure of human and bovine β2-glycoprotein I revealed by small-angle X-ray scattering. *J. Mol. Biol.* **321**, 85-97.

Hannan, J. P., Young, K. A., Guthridge, J. M., Asokan, R., Szakonyi, G., Chen, X. S. & Holers, V. M. (2005). Mutational analysis of the complement receptor type 2 (CR2/CD21)-C3d interaction reveals a putative charged SCR1 binding site for C3d. *J. Mol. Biol.* **346**, 845-858.

Hansen, J. C., Lebowitz, J. & Demeler, B. (1994). Analytical ultracentrifugation of complex macromolecular systems. *Biochemistry*, **33**, 13155-13163.

Harding, S. E. & Johnson, P. (1985). The concentration-dependence of macromolecular parameters. *Biochem. J.* **231**, 543-547.

Harris, L. J., Larson, S. B., Hasel, K. W. & McPherson, A. (1997). Refined structure of an intact IgG2a monoclonal antibody. *Biochemistry*, **36**, 1581-1597.

Hase, S., Kikuchi, N., Ikenaka, T. & Inoue, K. (1985). Structures of sugar chains of the third component of human complement. *J. Biochem*. **98**, 863-874.

Hasty, L. A. & Lyttle, C. R. (1992). Progesterone and RU486 regulation of uterine complement C3 after prior induction with estradiol. *Biol. Reprod*. **47**, 285-290.

Heenan, R. K., King, S. M., Turner, D. S. & Treadgold, J. R. (2006). SANS2D at the ISIS Second Target Station, Los Alamos National Laboratory Report, pp. 780-785.

Heinen, S., Hartmann, A., Lauer, N., Wiehl, U., Dahse, H. M., Schirmer, S., Gropp, K., Enghardt, T., Wallich, R., Hälbich, S., Mihlan, M., Schlötzer-Schrehardt, U., Zipfel, P. F. & Skerka, C. (2009). Factor H related protein 1 (CFHR-1) inhibits complement C5 convertase activity and terminal complex formation. *Blood,* **114**, 2439-2447.

Hebell, T., Ahearn, J. M. & Fearon, D. T. (1991). Suppression of the immune response by a soluble complement receptor of B lymphocytes. *Science*, **254**, 102-105.

Hellwage, J., Jokiranta, T. S., Friese, M. A., Wolk, T. U., Kampen, E., Zipfel, P. F. & Meri, S. (2002). Complement C3b/C3d and cell surface polyanions are recognized by overlapping binding sites on the most carboxyl-terminal domain of complement factor H. *J. Immunol.* **169**, 6935-6944.

Herbert, A. P., Uhrin, D., Lyon, M., Pangburn, M. K. & Barlow, P. N. (2006). Disease-associated sequence variations congregate in a polyanion recognition patch on human factor H revealed in three-dimensional structure. *J. Biol. Chem*. **281**,16512-16520.

Hirani, S., Lambris, J. D. & Müller-Eberhard, H. J. (1986). Structural analysis of the asparagine-linked oligosaccharides of human complement component C3. *Biochem. J.* **233**, 613-616.

Hourcade, D., Holers, V. M. & Atkinson, J. P. (1989). The regulators of complement activation (RCA) gene cluster. *Adv. Immunol*. **45**, 381-416.

Huang, J. L. & Lin, C. Y. (1994). A hereditary C3 deficiency due to aberrant splicing of exon 10. *Clin. Immunol. Immunopathol*. **73**, 267-273.

Hugli, T. E. (1975). Human anaphylatoxin (C3a) from the third component of complement. Primary structure. *J. Biol. Chem*. **250**, 8293-8301.

Hugli, T. E. (1981). The structural basis for anaphylatoxin and chemotactic functions of C3a, C4a, and C5a. *Crit. Rev. Immunol.* **1**, 321-366.

Ikeda, K., Sannoh, T., Kawasaki, N., Kawasaki, T. & Yamashina, I. (1987). Serum lectin with known structure activates complement through the classical pathway. *J. Biol. Chem.* **262**, 7451-7454.

Isenman, D. E., Leung, E., Mackay, J. D., Bagby, S. & van den Elsen, J. M. (2010). Mutational analyses reveal that the staphylococcal immune evasion molecule Sbi and complement receptor 2 (CR2) share overlapping contact residues on C3d: implications for the controversy regarding the CR2/C3d cocrystal structure. *J. Immunol*. **184**, 1946-1955.

Jakubik, J. J., Saifuddin, M., Takefman, D. M. & Spear, G. T. (1999). B lymphocytes in lymph nodes and peripheral blood are important for binding immune complexes containing HIV-1. *Immunology*, **96**, 612-619.

Janeway, C. A., Travers, P., Walport, M. & Schlomchik, M. J. (2005). Immunobiology: the immune system in health and disease. Garland Science Publishing, New York. pp. 55-73.

Janssen, B. J., Huizinga, E. G., Raaijmakers, H. C., Roos, A., Daha, M. R., Nilsson-Ekdahl, K., Nilsson, B. & Gros, P. (2005). Structures of complement component C3 provide insights into the function and evolution of immunity. *Nature*, **22**, 505-511.

Janssen, B. J., Christodoulidou, A., McCarthy, A., Lambris, J. D. & Gros, P. (2006). Structure of C3b reveals conformational changes that underlie complement activity. *Nature,* **444**, 213-216.

Janssen, B. J. C., Read, R. J., Brűnger, A. T. & Gros, P. (2007). Crystallographic evidence for deviating C3b structure? *Nature*, **448**, E1-E2.

Janssen, B. J., Halff, E. F., Lambris, J. D. & Gros, P. (2007). Structure of compstatin in complex with complement component C3c reveals a new mechanism of complement inhibition. *J. Biol. Chem.* **282**, 29241-29247.

Ji, Y. H., Matsushita, M., Okada, H., Fujita, T. & Kawakami, M. (1988). The C4 and C2 but not C1 components of complement are responsible for the complement activation triggered by the Ra-reactive factor. *J. Immunol.* **141**, 4271-4275.

Jokiranta, T. S., Hellwage, J., Koistinen, V., Zipfel, P. F. & Meri, S. (2000). Each of the three binding sites on complement factor H interacts with a distinct site on C3b. *J. Biol. Chem.* **275**, 27657-27662.

Jokiranta, T. S., Westin, J., Nilsson, U. R., Nilsson, B., Hellwage, J., Löfås, S., Gordon, D. L., Ekdahl, K. N. & Meri, S. (2001). Complement C3b interactions studied with surface plasmon resonance technique. *Int. Immunopharmacol.* **1**, 495-506.

Jokiranta, T. S., Jaakola, V. P., Lehtinen, M. J., Pärepalo, M., Meri, S. & Goldman, A. (2006). Structure of complement factor H carboxyl-terminus reveals molecular basis of atypical haemolytic uremic syndrome. *EMBO J.* **25**, 1784-1794.

Jones, J. F., Shurin, S., Abramowsky, C., Tubbs, R. R., Sciotto, C. G., Wahl, R., Sands, J. M., Gottman, D., Katz, B. Z. & Sklar, J. (1988). T-cell lymphomas containin Epstein-Barr viral DNA in patients with chronic Epstein-Barr virus infections. *N. Engl. J. Med.* **318**, 733-741.

Józsi, M., Heinen, S., Hartmann, A., Ostrowicz, C. W., Halbich, S., Richter, H., Kunert, A., Licht, C., Saunders, R. E., Perkins, S. J., Zipfel, P. F. & Skerka, C. (2006). Factor H and atypical hemolytic uremic syndrome: mutations in the C-terminus cause structural changes and defective recognition functions. *J. Am. Soc. Nephrol.* **17**, 170-177.

Kawasaki, T., Etoh, R. & Yamashina, I. (1978). Isolation and characterization of a mannan-binding protein from rabbit liver. *Biochem. Biophys. Res. Commun.* **81**, 1018-1024.

Kawasaki, N., Kawasaki, T. & Yamashina, I. (1983). Isolation and characterization of a mannan-binding protein from human serum. *J. Biochem. (Tokyo)*, **94**, 937-947.

Koch, M. H., Vachette, P. & Svergun, D. I.(2003). Small-angle scattering: a view on the properties, structures and structural changes of biological macromolecules in solution. *Q. Rev. Biophys*. **36**, 147-227.

Kovacs, J. M., Hannan, J. P., Eisenmesser, E. Z. & Holers, V. M. (2009). Mapping of the C3d ligand binding site on complement receptor 2 (CR2/CD21) using nuclear magnetic resonance and chemical shift analysis. *J. Biol. Chem.* **284**, 9513–9520.

Kühn, S. & Zipfel, P. F. (1996). Mapping of the domains required for decay acceleration activity of the human factor H-like protein 1 and factor H. *Eur. J. Immunol*. **26**, 2383-2397.

Lamm, O. (1929). Die Differentialgleichung der Ultrazentrifugierung. *Ark. Mat. Astr. Fys.* **21B**, 1-4.

Lambris, J. D., Avila, D., Becherer, J. D. & Müller-Eberhard, H. J. (1988). A discontinuous Factor H binding site in the third component of complement as delineated by synthetic peptides. *J. Biol. Chem.* **263**, 12147-12150.

Latham, M. P., Brown, D. J., McCallum, S. A. & Pardi, A. (2005). NMR methods for studying the structure and dynamics of RNA. *Chembiochem*. **6**, 1492-1505.

Laue, T. M., Shah, B. D., Ridgeway, T. M. & Pelletier, S. L. (1992). Computer-aided interpretation of analytical sedimentation data for proteins. *In* Analytical Ultracentrifugation in Biochemistry and Polymer Science. S. E. Harding, A. J. Rowe and J. C. Horton (eds.), The Royal Society of Chemistry, Cambridge, U.K. pp. 90-125.

Law, S. K. A. & Reid, K. B. M. (1995). Complement (Second edition). IRL Press, Oxford.

Lehtinen, M. J., Rops, A. L., Isenman, D. E., van der Vlag, J. & Jokiranta, T. S. (2009). Mutations of factor H impair regulation of surface-bound C3b by three mechanisms in atypical hemolytic uremic syndrome. *J. Biol. Chem*. **284**, 15650-15658.

Levy, E., Ambrus, J., Kahl, L., Molina, H., Tung, K. & Holers, V. M. (1992). T lymphocyte expression of complement receptor 2 (CR2/CD21): a role in adhesive cell-cell interactions and dysregulation in a patient with systemic lupus erythematosus (SLE). *Clin. Exp. Immunol.* **90**, 235-244.

Li, K., Okemefuna, A. I., Gor, J., Hannan, J. P., Asokan, R., Holers, V. M. & Perkins, S. J. (2008). Solution structure of the complex formed between human complement C3d and full length complement receptor Type 2. *J. Mol. Biol.* **384**, 137-150.

Li, K., Gor, J. & Perkins, S. J. (2010). Self-association and domain rearrangements between complement C3 and C3u provide insight into the activation mechanism of C3. *Biochem. J.* **431**, 63-72.

Liu, Y. J., Cairns, J. A., Holder, M. J., Abbot, S. D., Jansen, K. U., Bonnefoy, J.-Y., Gordon, J. & Maclennan, I. C. (1991). Recombinant 25-kDa CD23 and interleukin 1 alpha promtote the survival of germinal center B cells: evidence for bifurcation in the development of centrocytes rescued from apoptosis. *Eur. J. Immunol.* **21**, 1107-1114.

Lowell, C. A., Klickstein, L. B., Carter, R. H., Mitchell, J. A., Fearon, D. T. & Ahearn, J. M. (1989). Mapping of the Epstein-Barr virus and C3dg binding sites to a common domain on complement receptor type 2. *J. Exp. Med.* **170**, 1931-1946.

Makar, K. W., Pham, C. T. N., Dehoff, M. H., O'Connor, S. M., Jacobi, S. M. & Holers, V. M. (1998). An intronic silencer regulates B lymphocyte cell- and stage-specific expression of the human complement receptor type 2 (CR2, CD21) gene. *J. Immunol.* **160**, 1268-1278.

Manuelian, T., Hellwage, J., Meri, S., Caprioli, J., Noris, M., Heinen, S., Jozsi, M., Neumann, H. P., Remuzzi, G. & Zipfel, P. F. (2003). Mutations in factor H reduce binding affinity to C3b and heparin and surface attachment to endothelial cells in hemolytic uremic syndrome. *J. Clin. Invest*. **111**, 1181-1190.

Martin, D. R., Yuryev, A., Kalli, K. R., Fearon, D. T. & Ahern, J. M. (1991). Determination of the structural basis for selective binding of Epstein-Barr virus to human complement receptor type 2. *J. Exp. Med*. **174**, 1299-1311

Matsushita, M. & Fujita, T. (1992) Activation of the classical complement pathway by mannose-binding protein in association with a novel C1s-like serine protease, *J. Exp. Med.* **176,** 1497-1502.

McRorie, D. K. & Voelker, P. J. (1993). Self-Associating Systems in the Analytical Ultracentrifuge. Beckman Instruments, Fullerton, California.

Miki, K., Ogata, S., Misumi, Y. & Ikehara, Y. (1986). Carbohydrate structures of the third component of rat complement. Presence of both high-mannose and complex type oligosaccharide chains. *Biochem. J*. **240**, 691-698.

Molenaar, J. L., Muller, M. & Pondman, K. W. (1973). A new preparative method for isolation of human C3 with affinity chromatography. *J. Immunol.* **110**, 1570-1574.

Moore, M. D., Cooper, N. R., Tack, B. F. & Nemerow, G. R. (1987). Molecular cloning of the cDNA encoding the Epstein-Barr virus/C3d receptor (complement receptor type 2) of human B lymphocytes. *Proc. Natl. Acad. Sci. U. S. A*. **84**, 9194-9198.

Moore, M. D., DiScipio, R. G., Cooper, N. R. & Nemerow, G. R. (1989). Hydrodynamic, electron microscopic and ligand-binding analysis of the Epstein-Barr virus/C3dg receptor (CR2). *J. Biol. Chem.* **264**, 20576-20582.

Moore, M. D., Cannon, M. J., Sewall, A., Finlayson, M., Okimoto, M. & Nemerow, G. R. (1991). Inhibition of Epstein-Barr virus infection in vitro and in vivo by soluble CR2 (CD21) containing two short consensus repeats. *J. Virol.* **65**, 3559-3565.

Morgan, B. P. & Harris, C. L. (1999). Complement Regulatory Proteins. Academic Press, San Diego.

Morikis, D. & Lambris, J. D. (2004). Physical methods for structure, dynamics and binding in immunological research. *Trends*. *Immunol*. 25, 700-707. Review.

Morikis, D. & Lambris, J. (2005). Structural biology of the complement system. CRC Press, Boca Raton, FL. pp. 1-18.

Morley, B. J. & Walport, M. J. (2000). The Complement Facts Book. Academic Press, London.

Morris, K. M., Aden, D. P., Knowles, B. B. & Colten, H. R. (1982). Complement biosynthesis by the human hepatoma-derived cell line HepG2. *J. Clin. Invest*. **70**, 906-913.

Müller-Eberhard, H. J. (1968). Chemistry and Reaction Mechanisms of Complement. *Adv. Immunol*. **8**, 1-80.

Nagar, B., Jones, R. G., Diefenbach, R. J., Isenman, D. E. & Rini, J. M. (1998). X-ray crystal structure of C3d: a C3 fragment and ligand for complement receptor 2. *Science*, **280**, 1277-1281.

Nan, R., Gor, J., Lengyel, I. & Perkins, S. J. (2008a). Uncontrolled zinc- and copper-induced oligomerisation of the human complement regulator factor H and its possible implications for function and disease. *J. Mol. Biol*. **384**, 1341-1352.

Nan, R., Gor, J. & Perkins, S. J. (2008b). Implications of the progressive self-association of wild-type human Factor H for complement regulation and disease. *J. Mol. Biol.* **375**, 891-900.

Nan, R., Ward, G., Gavigan, L., Miller, A., Gor, J., McKay, A. R., Lengyel, I. & Perkins, S. J. (2010). Self-association and folded-back solution structures of the wild-type Tyr402 and the disease-related His402 allotypes of complement Factor H. Submitted.

Narayanan, T., Diat, O. & Bosecke, P. (2001). SAXS and USAXS on the high brilliance beamline at the ESRF. *Nucl. Instrum. Methods Phys. Res. A*. **467**, 1005-1009.

Narayanan, T., Sztucki, M., Belina, G. & Pignon, F. (2006). Microstructure and rheology near an attractive colloidal glass transition. *Phys. Rev. Lett*. **96**, 1-4.

Nelson, R. A. Jr., Jensen, J., Gigli, I. & Tamura, N. (1966). Methods for the separation, purification and measurement of nine components of haemolytic complement in guinea-pig serum. *Immunochemistry*, **3**, 111-135.

Nicol, P. A. & Lachmann, P. J. (1973). The alternate pathway of complement activation. The role of C3 and its inactivator (KAF). *Immunology*, **24**, 259-75.

Nishida, N., Walz, T. & Springer, T. A. (2006). Structural transitions of complement component C3 and its activation products. *Proc. Natl. Acad. Sci. U. S. A*. **103**, 19737-19742.

Nuttall, G. (1888). Experimente über die bacterienfeindlichen einfluesse des thierischen körpers. *Z. Hyg. Infektionskr.* **4**, 353-394.

Okemefuna, A. I., Gilbert, H. E., Griggs, K. M., Ormsby, R. J., Gordon, D. L., Perkins, S. J. (2008). The regulatory SCR-1/5 and cell surface-binding SCR-16/20 fragments of factor H reveal partially folded-back solution structures and different self-associative properties. *J. Mol. Biol.* **375**, 80-101.

Okemefuna, A. I. (2009). Complement Factor H: Solution structures and interactions with ligands. Ph.D thesis, University College London.

Okemefuna, A. I., Nan, R., Gor, J. & Perkins, S. J. (2009a). Electrostatic interactions contribute to the folded-back conformation of wild-type human Factor H. *J. Mol. Biol.* **391**, 98-118.

Okemefuna, A. I., Li, K., Nan, R., Ormsby, R. J., Sadlon, T., Gordon, D. L. & Perkins, S. J. (2009b). Multimeric interactions between complement Factor H and its C3d ligand provide new insight on complement regulation. *J. Mol. Biol.* **391**, 119-135.

Okemefuna, A. I., Gor, J., Sadlon, T., Adamson, P., Gordon, D. L. & Perkins, S. J. (2010a). A highly-looped central domain arrangement for complement Factor H clarifies its regulatory mechanism. In preparation.

Okemefuna, A. I., Stach, L., Rana, S., Ziai Buetas, A. J., Gor, J. & Perkins, S. J. (2010b). C-reactive protein exists in an NaCl concentration dependent pentamer-decamer equilibrium in physiological buffer. *J. Biol. Chem*. **285**, 1041-1052.

Okemefuna, A. I., Nan, R., Miller, A. Gor, J. & Perkins, S. J. (2010c). Complement Factor H binds at two independent sites to C-reactive protein in acute-phase concentrations. *J. Biol. Chem.* **285**, 1053-1065.

Oppermann, M., Manuelian, T., Jozsi, M., Brandt, E., Jokiranta, T. S., Heinen, S., Meri, S., Skerka, C., Gotze, O. & Zipfel, P. F. (2006). The C-terminus of complement regulator Factor H mediates target recognition: evidence for a compact conformation of the native protein. *Clin. Exp. Immunol.* **144**, 342-352.

Pangburn, M. K., Schreiber, R. D. & Müller-Eberhard, H. J. (1977). Human complement C3b inactivator: isolation, characterization, and demonstration of an absolute requirement for the serum protein beta1H for cleavage of C3b and C4b in solution. *J. Exp. Med.* **146**, 257-270.

Pangburn, M. K. & Müller-Eberhard, H. J. (1978). Complement C3 convertase: cell surface restriction of beta1H control and generation of restriction on neuraminidase-treated cells. *Proc. Natl. Acad. Sci. U. S. A*. **75**, 2416-2420.

Pangburn, M. K., Atkinson, M. A. L. & Meri, S. (1991). Localization of the heparin binding site on complement factor H. *J. Biol. Chem*. **266**, 16847-16853.

Pangburn, M. K., Rawal, N., Cortes, C., Alam, M. N., Ferreira, V. P. & Atkinson, M. A. (2009). Polyanion-induced self-association of complement factor H. *J. Immunol*. **182**, 1061-1068.

Panine, P., Finet, S., Weiss, T. M. & Narayanan, T. (2006). Probing fast kinetics in complex fluids by combined rapid mixing and small-angle X-ray scattering. *Adv. Colloid. Interface. Sci.* **127**, 9-18.

Pâques, E. P. (1980). Purification and partial characterization of the third component of the complement system from porcine serum (C3) and of a crystallizable degradation product of the fourth component of the complement system from human serum (C4). Study of the tryptic digestion products of human and porcine C3 and of human C4. *Hoppe-Seyler's Z. Physiol. Chem.* **361**, 445-456

Pepys, M. B. (1972). Role of complement in induction of the allergic response. *Nat. New. Biol.* **297**, 157.

Perkins, S. J. & Weiss, H. (1983). Low resolution structural studies of mitochondrial ubiquinol-cytochrome c reductase in detergent solutions by neutron scattering. *J. Mol. Biol.* **168**, 847–866.

Perkins, S. J., Kerckaert, J. P. & Loucheux-Lefebvre, M. H. (1985) The shapes of biantennary and tri-tetraantennary α_1 acid glycoprotein by small angle neutron and X-ray scattering. *Eur. J. Biochem.* **147**, 525-531.

Perkins, S. J. & Sim, R. B. (1986). Molecular modelling of human complement component C3 and its fragments by solution scattering. *Eur. J. Biochem*. **157**, 155-168.

Perkins, S. J. (1986). Protein volumes and hydration effects. The calculations of partial specific volumes, neutron scattering matchpoints and 280-nm absorption coefficients for proteins and glycoproteins from amino acid sequences. *Eur. J. Biochem*. **157**, 169-180.

Perkins, S. J. (1988). Structural studies of proteins by high-flux X-ray and neutron solution scattering. *Biochem. J.* **254**, 313-327.

Perkins, S. J., Nealis, A. S. & Sim, R. B. (1990a) Molecular modelling for human complement component C4 and its fragments by X-ray and neutron solution scattering. *Biochemistry,* **29**, 1167-1175.

Perkins, S. J., Smith, K. F., Nealis, A. S., Lachmann, P. J. & Harrison, R. A. (1990b) Structural homologies of component C5 of human complement with components C3 and C4 by neutron scattering. *Biochemistry*, **29**, 1175-1180.

Perkins, S. J., Nealis, A. S. & Sim, R. B. (1991). Oligomeric domain structure of human complement factor H by X-ray and neutron solution scattering. *Biochemistry,* **30**, 2847-2857.

Perkins, S. J. (2001a). Protein-Ligand Interactions: Hydrodynamics and Calorimetry. (Harding S. E. & Chowdhry, B. Z., eds). pp. 223-262, Oxford University Press, Oxford.

Perkins, S. J. (2001b). X-ray and neutron scattering analyses of hydration shells: a molecular interpretation based on sequence predictions and modelling fits. *Biophys. Chem.* **93**, 129-139.

Perkins, S. J., Gilbert, H. E., Aslam, M., Hannan, J. P., Holers, V. M. & Goodship, T. H. J. (2002). Solution structures of complement components by X-ray and neutron scattering and analytical ultracentrifugation. *Biochem. Soc. Transact.* **30**, 996-1001.

Perkins, S. J., Gilbert, H. E., Lee, Y. C., Sun, Z. & Furtado, P. B. (2005). Relating small angle scattering and analytical ultracentrifugation in multidomain proteins. In *Modern Analytical Ultracentrifugation: Techniques and Methods* (Scott, D. J., Harding, S. E. & Rowe, A. J., eds), pp. 291-319. Royal Society of Chemistry, London.

Perkins, S. J., Okemefuna, A. I., Fernando, A. N., Bonner, A., Gilbert, H. E. & Furtado, P. B. (2008). X-ray and neutron scattering data and their constrained molecular modelling. *Meth. Cell. Biol.* **84**, 375-423.

Perkins, S. J., Okemefuna, A. I., Nan, R., Li, K. & Bonner, A. (2009). Constrained solution scattering modelling of human antibodies and complement proteins reveals novel biological insights. *J. Roy. Soc. Interface*. **6**, S679-S696.

Perkins, S. J., Okemefuna, A. I., & Nan, R. (2010a). Unravelling protein-protein interactions between complement factor H and C-reactive protein by a multidisciplinary strategy. *Biochem. Soc. Transact*. **38**, 894-900.

Perkins, S. J., Nan, R., Okemefuna, A. I., Li, K., Khan, S. & Miller, A. (2010b). Multiple interactions of complement factor H with its ligands in solution: a progress report. Current Topics on Complement and Eye Diseases (Ed. J.D. Lambris and A. Adamis). *Adv. Exp. Med. Biol*. **703**, 25-47.

Philo, J. S. (2006). Improved methods for fitting sedimentation coefficient distributions derived by time-derivative techniques. *Anal. Biochem.* **354**, 238-246.

Pillemer, L., Ecker, E. E., Oncley, J. L. & Cohn, E. J. (1941). The preparation and physiochemical characterization of the serum protein components of complement. *J. Exp. Med*. **74**, 297-308.

Pillemer, L., Blum, L., Lepow, I. H., Ross, O. A., Todd, E. W. & Wardlaw, A. C. (1954). The properdin system and immunity. I. Demonstration and isolation of a new serum protein, properdin, and its role in immune phenomena. *Science*, **120**, 279-285.

Preissner, K. T. & Seiffert, D. (1998). Role of vitronectin and its receptors in haemostasis and vascular remodeling. *Thromb. Res*. **89**, 1-21.

Prodinger, W. M., Hellwage, J., Spruth, M., Dierich, M. P. & Zipfel, P. F. (1998). The C-terminus of factor H: monoclonal antibodies inhibit heparin binding and identify epitopes common to factor H and factor H-related proteins. *Biochem. J*. **331**, 41-47.

Prota, A. E., Sage, D. R., Stehle, T. & Fingeroth, J. D. (2002). The crystal structure of human CD21: Implications for Epstein-Barr virus and C3d binding. *Proc. Natl. Acad. Sci. U. S. A*. **99**, 10641-10646.

Pusey, P. N. & Tough, R J. A. (1985). Particle interactions. In *Dynamic Light Scattering: Applications of Photon Correlation Spectroscopy*. (Pecora, R. eds.). Plenum Press, New York, NY. pp. 85-179.

Putnam, C. D., Hammel, M., Hura, G. L. & Tainer, J. A. (2007). X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Q. Rev. Biophys.* **40**, 191-285.

Puy, L. A., Kuivanen, P. C. & DeSombre, E. R. (1993). Immunohistochemical localization of the oestrogen-responsive 110 kDa and 74 kDa polypeptides and complement component C3 in the rat genital tract after oestrogen treatment and during the oestrous cycle. *J. Reprod. Fertil*. **99**, 385-394.

Ralston, G. (1993). Introduction to Analytical Ultracentrifugation, Beckman Instruments, Palo Alto.

Ram, S., Sharma, A. K., Simpson, S. D., Gulati, S., McQuillen, D. P., Pangburn, M. K. & Rice, P. A. (1998a). A novel sialic acid binding site on factor H mediates serum resistance of sialylated Neisseria gonorrhoeae. *J. Exp. Med.* **187**, 743-52.

Ram, S., McQuillen, D. P., Gulati, S., Elkins, C., Pangburn, M. K. & Rice, P. A. (1998b). Binding of complement factor H to loop 5 of porin protein 1A: a molecular mechanism of serum resistance of nonsialylated Neisseria gonorrhoeae. *J. Exp. Med*. **188**, 671-680.

Raschke, T. M. (2006). Water structure and interactions with protein surfaces. *Curr. Opin. Struct. Biol*. **16**, 152-159.

Reid, K. B., Bentley, D. R., Campbell, R. D., Chung, L. P., Sim, R. B., Kristensen, T. & Tack, B. F. (1986). Complement system proteins which interact with C3b or C4b A superfamily of structurally related proteins. *Immunol. Today.* **7**, 230-234.

Reid, K. B. & Porter, R. R. (1981). The proteolytic activation systems of complement. *Annu. Rev. Biochem.* **50**, 433-464.

Reynes, M., Aubert, J. P., Cohen, J. H. M., Audouin, J., Tricottet, V., Diebold, J. & Kazatchkine, M. D. (1985). Human follicular dendritic cells express CR1, CR2, and CR3 complement receptor antigens. *J. Immunol.* **135**, 2687-2694.

Rey-Millet, C. A., Villiers, C. L., Gabert, F. M., Chesne, S. & Colomb, M. G. (1994). C3b covalently associated to tetanus toxin modulates TT processing and presentation by U937 cells. *Mol. Immunol*. **31**, 1321-1327.

Richards, E. G., Teller, D. C., Hoagland, V. D. J., Haschemeyer, R. H. & Schachman, H. K (1971). Alignment of Schlieren and Rayleigh Optical Systems in the Ultracentrifuge. II. A General Procedure. *Anal. Biochem*. **41**, 215-247.

Ripoche, J., Day, A. J., Harris, T. J. & Sim, R. B. (1988). The complete amino acid sequence of human complement factor H. *Biochem. J*. **249**, 593-602.

Rodriguez de Cordoba, S., Lublin, D. M., Rubinstein, P. & Atkinson, J. P. (1985). Human genes for three complement components that regulate the activation of C3 are tightly linked. *J. Exp. Med*. **161**, 1189-1195.

Rodríguez de Córdoba, S., Díaz-Guillén, M. A., Heine-Suñer, D. (1999). An integrated map of the human regulator of complement activation (RCA) gene cluster on 1q32. *Mol. Immunol.* **36**, 803-808.

Rooijakkers, S, H., Wu, J., Ruyken, M., van Domselaar, R., Planken, K. L., Tzekou, A., Ricklin, D., Lambris, J. D., Janssen, B. J., van Strijp, J. A. & Gros, P. (2009). Structural and functional implications of the alternative complement pathway C3 convertase stabilized by a staphylococcal inhibitor. *Nat. Immunol*. **10**, 721-727.

Sánchez-Corral, P., Pérez-Caballero, D., Huarte, O., Simckes, A. M., Goicoechea, E., López-Trascasa, M. & de Córdoba, S. R. (2002). Structural and functional characterization of factor H mutations associated with atypical hemolytic uremic syndrome. *Am. J. Hum. Genet*. **71**, 1285-1295.

Saphire, E. O., Parren, P. W., Pantophlet, R., Zwick, M. B., Morris, G. M., Rudd, P. M., Dwek, R. A., Stanfield, R. L., Burton, D. R. & Wilson, I. A. (2001). Crystal structure of a neutralizing human IgG against HIV-1: a template for vaccine design. *Science*, **293**, 1155-1159.

Sarrias, M. R., Franchini, S., Canziani, G., Argyropoulos, E., Moore, W. T. Sahu, A. & Lambris, J. D. (2001). Kinetic analysis of the interactions of complement receptor type 2 (CR2, CD21) with its ligands C3d, iC3b, and the EBV glycoprotein gp350/220. *J. Immunol.* **167**, 1490-1499.

Saunders, R. E., Goodship, T. H., Zipfel, P. F. & Perkins, S. J. (2006). An interactive web database of factor H-associated hemolytic uremic syndrome mutations: insights into the structural consequences of disease-associated mutations. *Hum. Mutat.* **27**, 21-30.

Saunders, R. E., Abarrategui-Garrido, C., Frémeaux-Bacchi, V., Goicoechea de Jorge, E., Goodship, T. H. J., López Trascasa, M., Noris, M., Ponce Castro, I. M., Remuzzi, G., Rodríguez de Córdoba, S., Sánchez-Corral, P., Skerka, C., Zipfel, P. F. & Perkins, S. J. (2007). The interactive Factor H - atypical Haemolytic Uraemic Syndrome mutation database and website: Update and integration of Membrane

Cofactor Protein and Factor I mutations with structural models. *Human Mutation,* **28**, 222-234.

Schachman, H. K. (1959). Ultracentrifugation in Biochemistry. Academic Press, New York.

Schmidt, C. Q., Herbert, A. P., Kavanagh, D., Gandy, C., Fenton, C. J., Blaum, B. S., Lyon, M., Uhrin, D. & Barlow, P. N. (2008). A new map of glycosaminoglycan and C3b binding sites on Factor H. *J. Immunol*. **181**, 2610-2619.

Schuck, P. (1998). Sedimentation analysis of non-interacting and self-associating solutes using numerical solutions to the Lamm equation. *Biophys. J*. **75**, 1503-1512.

Schuck, P. (2000). Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and Lamm equation modeling. *Biophys. J*. **78**, 1606-1619.

Schuck, P. (2003). On the analysis of protein self-association by sedimentation velocity analytical ultracentrifugation. *Anal. Biochem.* **320**, 104-124.

Schuck, P. (2007). Protein Interactions: Biophysical Approaches for the Study of Complex Reversible Systems, Springer U. S. pp. 469-518.

Schuster, M. C., Ricklin, D., Papp, K., Molnar, K. S., Coales, S. J., Hamuro, Y., Sfyroera, G., Chen, H., Winters, M. S. & Lambris, J. D. (2008). Dynamic structural changes during complement C3 activation analyzed by hydrogen/deuterium exchange mass spectrometry. *Mol. Immunol*. **45**, 3142-3151.

Schwaeble, W., Zwirner, J., Schulz, T. F., Linke, R. P., Dierich, M. P. & Weiss, E. H. (1987). Human complement factor H: expression of an additional truncated gene product of 43 kDa in human liver. *Eur. J. Immunol*. **17**, 1485-1489.

Scott, D. J. & Schuck, P. (2005). A brief introduction to the analytical ultracentrifugation of proteins for beginners. In *Analytical Ultracentrifugation: Techniques and Methods*. (Scott, D. J., Harding, S. E. & Rowe, A. J., eds). pp. 1-25. Royal Society of Chemistry, Cambridge.

Semenyuk, A. V. & Svergun, D. I. (1991). GNOM - a program package for small-angle scattering data-processing. *J. Appl. Crystallogr.* **24,** 537-540.

Serdyuk, I. N., Zaccai, N. R. & Zaccai, J. (2007). Methods in Molecular Biophysics: Structure, Dynamics, Function. Cambridge University Press, Cambridge.

Sharma, A. K. & Pangburn, M. K. (1996). Identification of three physically and functionally distinct binding sites for C3b in human complement factor H by deletion mutagenesis. *Proc. Natl. Acad. Sci. U. S. A*. **93**, 10996-11001.

Sibanda, B. L. & Thornton, J. M. (1985). Beta-hairpin families in globular proteins. *Nature*, **316**, 170-174.

Sim, E., Wood, A. B., Hsiung, L. M. & Sim, R. B. (1981). Pattern of degradation of human complement fragment C3b. *FEBS Lett.* **132**, 55-60.

Sim, R.B., Day, A. J., Moffatt, B. E. & Fontaine, M. (1993). Complement factor I and cofactors in control of complement system convertase enzymes, *Meth. Enzymol*. **223**, 13–35.

Singer, L., Colten, H. R. & Wetsel, R. A. (1994a). Complement C3 deficiency: human, animal, and experimental models. *Pathobiology*, **62**, 14-28.

Singer, L., Whitehead, W. T., Akama, H., Katz, Y., Fishelson, Z. & Wetsel, R. A. (1994b). Inherited human complement C3 deficiency. An amino acid substitution in the β-chain (ASP^{549} to ASN) impairs C3 secretion. *J. Biol. Chem*. **269**, 28494-28499.

Singer, L., Van Hee, M. L., Lokki, M. L., Kramer, J., Borzy, M. S. & Wetsel, R. A. (1996) Inherited complement C3 deficiency: reduced C3 mRNA and protein levels in a Laotian kindred. *Clin. Immunol. Immunopathol*. **81**, 244-252.

Soames, C. J. & Sim, R. B. (1997). Interaction between human complement components factor H, factor I and C3b. *Biochem. J*. **326**, 553-561.

Solovyova, A., Schuck, P., Costenaro, L. & Ebel, C. (2001). Non-ideality by sedimentation velocity of halophilic malate dehydrogenase in complex solvents. *Biophys. J.* **81**, 1868-1880.

Sottrup-Jensen, L., Stepanik, T. M., Kristensen, T., Lønblad, P. B., Jones, C. M., Wierzbicki, D. M., Magnusson, S., Domdey, H., Wetsel, R. A. & Lundwall, A. (1985). Common evolutionary origin of alpha 2-macroglobulin and complement components C3 and C4. *Proc. Natl. Acad. Sci. U. S. A*. **82**, 9-13.

Stafford, W. F. (1992a). Boundary analysis in sedimentation transport experiments: a procedure for obtaining sedimentation coefficient distributions using the time derivative of the concentration profile. *Anal. Biochem.* **203**, 295-301.

Stafford, W. F. (1992b). Methods for obtaining sedimentation coefficient distributions. In *Analytical Ultracentrifugation in Biochemistry and Polymer Science.* (Harding, S. E., Rowe, A. J. & Horton, J. C., eds), pp. 359–393. Royal Society of Chemistry, Cambridge.

Stafford, W. F. (1994a). Sedimentation boundary analysis of interacting systems: use of the apparent sedimentation coefficient distribution function. In *Modern Analytical Ultracentrifugation.* (Shuster, T. M., Laue, T. M., eds), pp. 119–137. Birkhauser, Boston.

Stafford, W. F. (1994b). Boundary analysis in sedimentation velocity experiments. *Methods Enzymol.* **240**, 478-501.

Stafford, W. F. (2000). Analysis of reversibly interacting macromolecular systems by time derivative sedimentation velocity. *Methods. Enzymol.* **323,** 302-325.

Stafford, W. F. & Sherwood, P. J. (2004). Analysis of heterologous interacting systems by sedimentation velocity: curve fitting algorithms for estimation of sedimentation coefficients, equilibrium and kinetic constants. *Biophys. Chem.* **108**, 231-243.

Stoiber, H., Pintér, C., Siccardi, A. G., Clivio, A. & Dierich, M. P. (1996). Efficient destruction of human immunodeficiency virus in human serum by inhibiting the protective action of complement factor H and decay accelerating factor (DAF, CD55). *J. Exp. Med*. **183**, 307-310.

Stoiber, H., Ammann, C., Spruth, M., Müllauer, B., Eberhart, A., Harris, C, L., Huber, C. G., Longhi, R., Falkensammer, B., Würzner, R. & Dierich, M. P. (2001). Enhancement of complement-mediated lysis by a peptide derived from SCR 13 of complement factor H. *Immunobiology*, **203**, 670-686.

Svedberg. T. & Nichols, J. B. (1923). Determination of size and distribution of size of particle by centrifugal methods. *J. Am. Chem. Soc.* **45,** 2910-2917.

Svergun, D. I. & Stuhrmann, H. B. (1991). New developments in direct shape determination from small-angle scattering, 1. Theory and model calculations. *Acta. Crystallogr.* **A47**, 736.

Svergun, D. I. (1992). Determination of the regularization parameter in indirect transform methods using perceptual criteria. *J. Appl. Cryst.* **25**, 495-503.

Svergun, D. I., Richard, S., Koch, M. H., Sayers, Z., Kuprin, S. & Zaccai, G. (1998). Protein hydration in solution: experimental observation by x-ray and neutron scattering. *Proc. Natl. Acad. Sci. U. S. A*. **95**, 2267-2272.

Svergun, D. I. & Koch, M. H. (2003). Small-angle scattering studies of biological macromolecules in solution. *Rep. Prog. Phys*. **66**, 1735-1782.

Szakonyi, G., Guthridge, J. M., Li, D., Young, K., Holers, V. M. & Chen, X. S. (2001). Structure of complement receptor 2 in complex with its C3d ligand. *Science,* **292**, 1725-1728.

Tack, B. F., Harrison, R. A., Janatova, J., Thomas, M. L. & Prahl, J. W. (1980). Evidence for presence of an internal thiolester bond in third component of human complement. *Proc. Natl. Acad. Sci. U. S. A*. **77**, 5764-5768.

Tanford, C. (1961). *Physical Chemistry of Macromolecules*. pp. 340. Wiley, New York.

Tanious, F. A., Nguyen, B. & Wilson, W. D. (2008). Biosensor-surface plasmon resonance methods for quantitative analysis of biomolecular interactions. *Meth. Cell Biol.* **84**, 53-77.

Tedder, T. F., Clement, L. T. & Cooper, M. D. (1984). Expression of C3d receptors during human B cell differentiation: Immunofluoresence analysis with the HB-5 monoclonal antibody. *J. Immunol.* **133**, 678-683.

Teller, D. C. (1973). Characterization of proteins by sedimentation equilibrium in the analytical ultracentrifuge. *Methods Enzymol*. **27**, 346-441.

Thomas, M. L., Janatova, J., Gray, W. R. & Tack, B. F. (1982). Third component of human complement: localization of the internal thiolester bond. *Proc. Natl Acad. Sci. U. S. A.* **79**, 1054-1058.

Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic. Acids. Res.* **22**, 4673-4680.

Thyphronitis, G., Kinoshita, T., Inoue, K., Schweinle, J. E., Tsokos, G. S., Metcalf, E. S., Finkelman, F. D. &. Balow, J. E. (1991). Modulation of complement receptors I and 2 suppress antibody response *in vivo*. *J. Immunol.* **147**, 224-230.

Tissières, A. & Watson, J. D. (1958). Ribonucleoprotein particles from *Escherichia coli*. *Nature,* **182,** 778-780.

Tolnay, M. & Tsokos, G. C. (1998). Complement receptor 2 in the regulation of the immune response. *Clin. Immunol. Immunopath.* **88**, 123-132.

Toothaker, L. E., Henjes, A. J. & Weis, J. J. (1989). Variability of CR2 gene products is due to alternative exon usage and different CR2 alleles. *J. Immunol.* **142**, 3668-3675.

Tooze, R. M., Doody, G. M. & Fearon, D. T. (1997). Counterregulation by the coreceptors CD19 and CD22 of MAP kinase activation by membrane immunoglobulin *Immunity*. **7**, 59-67.

Tsoukas, C. D. & Lambris, J. D. (1988). Expression of CR2/EBV receptors on human thymocytes detected by monoclonal antibodies. *Eur. J. Immunol.* **18**, 1299-1302.

Ueki, A., Sai, T., Oka, H., Tabata, M., Hosokawa, K. & Mochizuki, Y. (1987). Biosynthesis and secretion of the third component of complement by human endothelial cells in vitro. *Immunology*, **61**, 11- 14.

Van Holde, K. E. (1975). Sedimentation analysis of proteins, 3rd ed. In *The Proteins* (Neurath, H. & Hill, R. H. eds), pp. 225-291. Academic Press, New York.

Van Holde, K. E. (1985). Physical Biochemistry, 2nd ed. Prentice-Hall. Englewood Cliffs, New Jersey.

Van Holde, K. E. & Weischet, W. O. (1978). Boundary analysis of sedimentation velocity experiments with monodisperse and paucidisperse solutes. *Biopolymers,* **17**, 1387-1403.

Walport, M. J. (2001). Complement. *N. Engl. J. Med.* **344**, 1058-1066 & 1140-1144.

Weiler, J. M., Daha, M. R., Austen, K. F. & Fearon, D. T. (1976). Control of the amplification convertase of complement by the plasma protein beta1H. *Proc. Natl. Acad. Sci. U. S. A*. **73**, 3268-3272.

Weis, J. J., Tedder, T. F. & Fearon, D. T. (1984). Identification of a 145,000 M_r membrane protein as the C3d receptor (CR2) of human B lymphocytes. *Proc. Nat. Acad. Sci. U. S. A.* **81**, 881-885.

Weis, J. J. & Fearon, D. T. (1985). The identification of N-linked oligosaccharides on the human CR2/Epstein-Barr virus receptor and their function in receptor metabolism, plasma membrane expression, and ligand binding. *J. Biol. Chem.* **260**, 13824-13830.

Weis, J. H., Morton, C. C., Bruns, G. A. P., Weis, J. J., Klickstein, L. B., Wong, W. W. & Fearon, D. T. (1987). A complement receptor locus: genes encoding C3b/C4b receptor and C3d/Epstein-Barr virus receptor MAP to 1q32. *J. Immunol.* **138**, 312-315.

Weis, J. J., Toothaker, L. E., Smith, J. A., Weis, J. H. & Fearon, D. T. (1988). Structure of the human B lymphocyte receptor for C3d and the Epstein-Barr virus and relatedness to other members of the family of C3/C4 binding proteins. *J. Exp. Med.* **167**, 1047-1066.

Weiss, L. M., Movahed, L. A., Warnke, R. A. & Sklar, J. (1989). Detection of Epstein-Barr viral genomes in Reed-Sternberg cells of Hodgkin's disease. *N. Engl. J. Med.* **320**, 502-506.

Welch, T. R., Beischel, L. & Kleesattel, A. (1990). Functional consequences of the genetic polymorphism of the third component of complement. *J. Pediatr*. **16**, S92-S97.

Welch, T. R., Beischel, L. S. & Witte, D. P. (1993). Differential expression of complement C3 and C4 in the human kidney. *J. Clin. Invest*. **92**, 1451-1458.

Whaley, K. & Ruddy, S. (1976). Modulation of the alternative complement pathways by beta 1 H globulin. *J. Exp. Med.* **144**, 1147-1163.

Wiesmann, C., Katschke, K. J., Yin, J., Helmy, K. Y., Steffek, M., Fairbrother, W. J., McCallum, S. A., Embuscado, L., DeForge, L., Hass, P. E. & van Lookeren Campagne, M. (2006). Structure of C3b in complex with CRIg gives insights into regulation of complement activation. *Nature*, **444**, 217-220.

Wills, P. R. & Winzor D. J. (1992) thermodynamic non-ideality and sedimentation equilibrium. *In* Analytical Ultracentrifugation in Biochemistry and Polymer Science. (Harding, S. E., Rowe, A. J. & Horton, J. C. eds.), The Royal Society of Chemistry, Cambridge, U.K. pp. 311-330.

Willey, S. & Aasa-Chapman, M. M. (2008). Humoral immunity to HIV-1: neutralisation and antibody effector functions. *Trends Microbiol.* **16,** 596-604.

Winters, M. S., Spellman, D. S. & Lambris, J. D. (2005). Solvent accessibility of native and hydrolyzed human complement protein 3 analyzed by hydrogen/deuterium exchange and mass spectrometry. *J. Immunol.* **174**, 3469-3474.

Winzor,D.J., Carrington L. E. & Harding, S. E. (2004a). Limitations of the ultracentrifugal approach for measuring the effective net charge of a macroion. *Anal. Biochem*. **333**, 114-118.

Winzor D. J., Jones, S. & Harding, S. E. (2004b). Determination of protein charge by capillary zone electrophoresis. *Anal. Biochem*. **333**, 225-229.

Wu, J., Wu, Y. Q., Ricklin, D., Janssen, B. J., Lambris, J. D. & Gros, P. (2009). Structure of complement fragment C3b-factor H and implications for host protection by complement regulators. *Nat. Immunol*. **10**, 728-733.

Wüthrich, K. (1986). NMR of Proteins and Nucleic Acids. Wiley-Interscience, New York.

Xu, Y., Zhang, C., Jia, L., Wen, C., Liu, H., Wang, Y., Sun, Y., Huang, L., Zhou, Y. & Song, H. (2009). A novel approach to inhibit HIV-1 infection and enhance lysis of HIV by a targeted activator of complement. *Virol. J.* **6,** 123.

Yates, J. R., Sepp, T., Matharu, B. K., Khan, J. C., Thurlby, D. A., Shahid, H., Clayton, D. G., Hayward, C., Morgan, J., Wright, A. F., Armbrecht, A. M., Dhillon, B., Deary, I. J., Redmond, E., Bird, A.C., Moore, A. T.; & Genetic Factors in AMD Study Group. (2007). Complement C3 variant and the risk of age-related macular degeneration. *N. Engl. J. Med.* **357**, 553-561.

Young, L. S., Sixbey, J. W., Clark, D. & Rickinson, A. B. (1986). Epstein-Barr virus receptors on human pharyngeal epithelia. *Lancet*, **327**, 240-242.

Young, K. A., Chen, X. S., Holers, V. M. & Hannan, J. P. (2007). Isolating the Epstein-Barr virus gp350/220 binding site on complement receptor type 2 (CR2/CD21). *J. Biol. Chem*. **282**, 36614-36625.

Young, K. A., Herbert, A. P., Barlow, P. N., Holers, V. M. & Hannan, J. P. Molecular basis of the interaction between complement receptor type 2 (CR2/CD21) and Epstein-Barr virus glycoprotein gp350. (2008). *J. Virol*. **82**, 11217-11227.

Zanotti, G., Bassetto, A., Battistutta R., Stoppini, M., Folli C. & Berni, R. (2000). Structure at 1.44 Å resolution of an N-terminally truncated form of the rat serum complement C3d fragment. *Biochim. Biophys. Acta.* **1478,** 232–238.

Zhang, L., Mallik, B. & Morikis, D. (2007). Immunophysical exploration of C3d-CR2 (CCP1-2) interaction using molecular dynamics and electrostatics. *J. Mol. Biol*. **369**, 567-583.

Publications

Li, K., Okemefuna, A. I., Gor, J., Hannan, J. P., Asokan, R., Holers, V. M. & Perkins, S. J. (2008). Solution structure of the complex formed between human complement C3d and full-length complement receptor type 2. *J. Mol. Biol*. **384**, 137-150.

Perkins, S. J., Okemefuna, A. I., Nan, R., Li, K. & Bonner, A. (2009). Constrained solution scattering modelling of human antibodies and complement proteins reveals novel biological insights. *J. R. Soc. Interface*. **6, Suppl 5**, S679-S696.

Okemefuna, A. I., Li, K., Nan, R., Ormsby, R. J., Sadlon, T., Gordon, D. L. & Perkins, S. J. (2009). Multimeric interactions between complement factor H and its C3d ligand provide new insight on complement regulation. *J. Mol. Biol*. **391**, 119-135.

Li, K., Gor, J. & Perkins, S. J. (2010). Self-association and domain rearrangements between complement C3 and C3u provide insight into the activation mechanism of C3. *Biochem. J. Biochem. J*. 431, 63-72.

Li, K., Okemefuna, A. I., Gor, J. & Perkins, S. J. (2010). Solution structures of the complexes between complement C3u and Factor H reveals two independent C3u sites on Factor H: implications for complement regulation. *In preparation.*

Perkins, S. J., Nan, R., Okemefuna A. I., Li, K., Khan, S. & Miller, A. (2010). Multiple interactions of complement Factor H with its ligands in solution: a progress report. *Adv. Exp. Med. Biol.* **703**, 25-47.

Presentations and abstracts

Li, K. Solution structure of human complement C3. Third year Graduate Student Symposium, University College London, London, U.K. 23rd June 2009 (Oral).

Li, K., Gor, J., Hannan, J. P., Asokan, R., Holers, V. M. & Perkins, S. J. Solution structure of the complex between complement C3d and full length complement receptor Type 2. 11th European meeting on Complement in Human Disease, Cardiff, U.K. 9th Sept-11th Sept 2007 (Poster).

Li, K., Gor, J., Hannan, J. P., Asokan, R., Holers, V. M. & Perkins, S. J. Solution structure of the complex between complement C3d and full length complement receptor Type

2. 22nd International Complement workshop, Basel, Switzerland, 28th Sept – 2nd Oct 2008 (Poster).

Li, K., Gor, J. & Perkins, S. J. Self-association and domain rearrangements between complement C3 and C3u provide insight into the activation mechanism of C3. 23rd International Complement workshop, New York, U.S.A., 1st Aug-5th Aug, 2010 (Poster).

Li, K., Gor, J. & Perkins, S. J. Self-association and domain rearrangements between complement C3 and C3u provide insight into the activation mechanism of C3. 19th International AUC Conference, Nottingham, U. K., 12nd Sept -16th Sept, 2010 (Oral).

Li, K., Miller, A. & Perkins, S. J. Purification of C3. AKTA User Days Meeting, London, U. K., 24th Nov, 2010 (Oral).